



सत्यमेव जयते

INDIAN AGRICULTURAL  
RESEARCH INSTITUTE, NEW DELHI

I.A.R.I.6.

GIP NLK- 3 I.A.R.I. -10-5 55--15,000







THE ANNALS  
*of*  
MATHEMATICAL  
STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF  
MATHEMATICAL STATISTICS

VOLUME XI

1940

# THE ANNALS OF MATHEMATICAL STATISTICS

EDITED BY

S. S. WILKS, *Editor*

A. T. CRAIG

J. NEYMAN

WITH THE COÖPERATION OF

H. C. CARVER

R. A. FISHER

R. DE MISES

H. CRAMÉR

T. C. FRY

E. S. PEARSON

W. E. DEMING

H. HOTELLING

H. L. RIETZ

G. DARMOIS

W. A. SHEWHART

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, P. R. Rider, Washington University, St. Louis, Mo.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is \$4.00 per year. Single copies \$1.25. Back numbers are available at the following rates:

Vols. I-IV \$5.00 each. Single numbers \$1.50.

Vols. V to date \$4.00 each. Single numbers \$1.25.

## CONTENTS OF VOLUME XI

### ARTICLES

BROWN, GEORGE W. Reduction of a Certain Class of Statistical Hypotheses .....	254
COCHRAN, W. G. The Analysis of Variance when Experimental Errors follow the Poisson or Binomial Laws .....	335
COX, GERTRUDE M. Enumeration and Construction of Balanced Incomplete Block Configurations .....	72
CRAIG, CECIL C. The Product Seminvariants of the Mean and a Central Moment in Samples .....	177
DALY, JOSEPH F. On the Unbiased Character of Likelihood-Ratio Tests for Independence in Normal Systems .....	1
DANTZIG, GEORGE B. On the Non-Existence of Tests of "Student's" Hypothesis having Power Functions Independent of $\sigma$ .....	186
DEMING, W. EDWARDS AND STEPHAN, FREDERICK F. On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known .....	427
DODD, EDWARD L. The Substitutive Mean and Certain Subclasses of this General Mean .....	163
DRESSEL, PAUL L. Statistical Seminvariants and their Estimates with Particular Emphasis on their Relation to Algebraic Invariants .....	33
DWYER, P. S. The Cumulative Numbers and their Polynomials .....	66
FRIEDMAN, MILTON. A Comparison of Alternative Tests of Significance for the Problem of $m$ Rankings .....	86
GEIRINGER, HILDA. A Generalization of the Law of Large Numbers .....	393
HOEL, PAUL G. The Errors Involved in Evaluating Correlation Determinants .....	58
HOTELLING, HAROLD. The Selection of Variates for Use in Prediction with some Comments on the Problem of Nuisance Parameters .....	271
HOTELLING, HAROLD. The Teaching of Statistics .....	457
HSU, C. T. On Samples from a Normal Bivariate Population .....	410
JOHNSON, N. L. Parabolic Test for Linkage .....	227
KENDALL, M. G. Conditions for Uniqueness in the Problem of Moments ..	402
MADOW, WILLIAM G. Limiting Distributions of Quadratic and Bilinear Forms .....	125
MOOD, A. M. The Distribution Theory of Runs .....	367
PIERCE, JOSEPH A. A Study of a Universe of $n$ Finite Populations with Application to Moment-Function Adjustments for Grouped Data .....	311

REIERSØL, OLAV. A Method for Recurrent Computation of all of the Principal Minors of a Determinant and its Application in Confluence Analysis .....	193
SINGLETON, ROBERT R. A Method for Minimizing the Sum of Absolute Values of Deviations .....	301
STEPHAN, FREDERICK F. AND DEMING, W. EDWARDS. On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known .....	427
WALD, ABRAHAM. The Fitting of Straight Lines if Both Variables are Subject to Error .....	284
WALD, A. AND WOLFOWITZ, J. On a Test Whether Two Samples are from the same Population .....	147
WOLFOWITZ, J. AND WALD, A. On a Test Whether Two Samples are from the Same Population .....	147

## NOTES

BAKER, G. A. Comparison of Pearsonian Approximations with Exact Sampling Distributions of Means and Variances in Samples from Populations Composed of the Sums of Normal Populations .....	219
BROWN, A. W. A Note on the Use of a Pearson Type III Function in Renewal Theory .....	448
BLEICK, W. E. A Least Squares Accumulation Theorem .....	225
COHEN, A. C., JR. The Numerical Computation of the Product of Conjugate Imaginary Gamma Functions .....	213
COCHRAN, W. G. Note on an Approximate Formula for Significance Levels of $z$ .....	93
DEMING, W. EDWARDS. Discussion of Professor Hotelling's paper on the Teaching of Statistics .....	470
DIXON, W. J. A Criterion for Testing the Hypothesis that Two Samples are from the same Population .....	199
DWYER, P. S. Combinatorial Formulas for the $r$ th Standard Moment of the Sample Sum, of the Sample Mean, and of the Normal Curve ...	353
EVANS, W. D. Note on the Moments of a Binomially Distributed Variate.	106
FRANKEL, A. AND KULLBACK, S. A Simple Sampling Experiment on Confidence Intervals .....	209
JOHNSON, EVAN, JR. Estimates of Parameters by Means of Least Squares.	453
KIMBALL, BRADFORD F. Orthogonal Polynomials Applied to Least Square Fitting of Weighted Observations .....	348
KULLBACK, S. AND FRANKEL, A. A Simple Sampling Experiment on Confidence Intervals .....	209
MADOW, WILLIAM G. The Distribution of Quadratic Forms in Non-Central Normal Random Variables .....	100
MAUCHLY, JOHN W. Significance Test for Sphericity of a Normal $n$ -variate Distribution .....	204

NORRIS, NILAN. The Standard Errors of the Geometric Means and their Application to Index Numbers.....	445
OLDS, E. G. On a Method of Sampling.....	355
OLMSTEAD, P. S. Note on Theoretical and Observed Distributions of Repetitive Occurrences.....	363
WALD, ABRAHAM. A Note on the Analysis of Variance with Unequal Class Frequencies.....	96
WOODBURY, MAX A. Rank Correlation when there are Equal Variates....	358

## MISCELLANEOUS

Abstracts of Papers.....	110, 475
Constitution and By-Laws of the Institute.....	116
Directory of the Institute.....	120
Report of the Philadelphia Meeting of the Institute.....	108
Report of the Hanover Meeting of the Institute.....	473
Report of the War Preparedness Committee of the Institute.....	479
Resolutions of the Institute on the Teaching of Statistics.....	472



# ON THE UNBIASED CHARACTER OF LIKELIHOOD-RATIO TESTS FOR INDEPENDENCE IN NORMAL SYSTEMS

BY JOSEPH F. DALY

1. **Introduction.** In the statistical interpretation of experimental data, the basic assumption is, of course, that we are dealing with a sample from a statistical population, the elements of which are characterized by the values of a number of random variables  $x^1, \dots, x^k$ . But in many cases we are in a position to assume even more, namely, that the population has an elementary probability law  $f(x^1, \dots, x^k; \theta_1, \dots, \theta_h)$ , where the functional form of  $f(x, \theta)$  is definitely specified, although the parameters  $\theta_1, \dots, \theta_h$  are to be left free for the moment to have values corresponding to any point of a set  $\Omega$  in an  $h$ -dimensional space.

Under this assumption, the problem of obtaining from the data further information about the hypothetical distribution law  $f(x, \theta)$  is considerably simplified. For it is then equivalent to that of deciding whether or not the data support the hypothesis that the population values of the  $\theta$ 's correspond to a point in a certain subset  $\omega$  of  $\Omega$ . For example, we may have reason to believe that the population  $K$  has a distribution law of the form

$$f(x^1, x^2; a^1, a^2, A_{11}, A_{12}, A_{22}) = \frac{|A_{ij}|^{\frac{1}{2}}}{2\pi} e^{-\frac{1}{2} \sum_{i,j} A_{ij}(x^i - a^i)(x^j - a^j)}.$$

Here the set  $\Omega$  is composed of all parameter points  $(a^1, \dots, A_{22})$  for which the matrix  $\|A_{ij}\|$  ( $i, j = 1, 2$ ) is positive definite and for which  $-\infty < a^i < \infty$ . We may wish to decide, on the basis of  $N$  independent observations  $(x^1_\alpha, x^2_\alpha)$  drawn from  $K$ , whether  $A_{12}$  has the value zero for the population in question, without concerning ourselves at all about the values of the remaining parameters; in other words, we may wish to test the hypothesis  $H$  that the parameter point corresponding to  $K$  lies in that subset of  $\Omega$  for which  $A_{12} = 0$ . One way to test this hypothesis is to select some (measurable) function  $g(x)$  whose value can be determined from the data, say

$$g(x) = \frac{\sum_{\alpha=1}^N (x^1_\alpha - \bar{x}^1)(x^2_\alpha - \bar{x}^2)}{\left[ \sum_{\alpha=1}^N (x^1_\alpha - \bar{x}^1)^2 \right]^{\frac{1}{2}} \left[ \sum_{\alpha=1}^N (x^2_\alpha - \bar{x}^2)^2 \right]^{\frac{1}{2}}}.$$

Now  $g(x)$  is itself a random variable, so that it has a distribution law of its own when its constituent  $x$ 's are drawn from any particular population  $K$ . Suppose then we choose a set of values of  $g(x)$ , say  $S$ , such that the probability is only .05 that  $g(x)$  will lie in the set  $S$  when the  $x$ 's are drawn independently from a population  $K$  for which the above hypothesis  $H$  is true. Ordinarily we would



take  $S$  to be of the form  $|g(x)| \geq g_0$ , and the test would then reject  $H$  at the .05 probability level if the computed value of  $g(x)$  came out too large. But for all that has been said so far, we are perfectly free to choose a different critical region  $S$ , and even a different function  $g(x)$ . The essential elements of this type of test are then a critical region  $S$ , a function of the data  $g$ , and a probability level  $\epsilon$ , such that the probability is  $\epsilon = .05$ , say, that  $g \subset S$  when  $H$  is true; in employing the test we reject  $H$  at the given probability level whenever the sample value of  $g$  falls in the critical region.

By the very nature of the problem, any inferences we make from a sample are subject to possible error. In the kind of test under consideration, the only error we can commit, strictly speaking, is that of rejecting  $H$  when it is true (an error of Type I in the terminology of Neyman and Pearson [9]). The risk of such an error is thus known in advance; for if we use the test consistently at, say, the .05 level, we know that the probability is .05 that we shall be led to reject a given hypothesis when it is true. On the other hand, it is quite conceivable that the test may be even less likely to reject  $H$  when it is false, or more precisely, when the true  $\theta$ 's correspond to a point of  $\Omega$  which is not in  $\omega$ . In this event the test is said to be biased. Let us make this term more definite by proposing the following definitions:

**DEFINITION I.** *A test is said to be completely unbiased if it has the property that for any probability level  $\epsilon$  ( $0 < \epsilon < 1$ ) the probability of rejecting  $H$  is greater when the  $\theta$ 's correspond to a point of  $\Omega - \omega$  than when they correspond to a point of  $\omega$ .*

**DEFINITION II.** *A test is said to be locally unbiased if the set  $\Omega$  contains a neighborhood  $U$  of  $\omega$  such that for any probability level  $\epsilon$  ( $0 < \epsilon < 1$ ) the probability of rejecting  $H$  is greater when the parameter values correspond to a point of  $U - \omega$  than when they correspond to a point of  $\omega$ .*

It is the purpose of this paper to consider the question of bias in connection with the Neyman-Pearson method of likelihood ratios [8] as applied to the testing of what may well be called hypotheses of independence in multivariate normal populations. The likelihood ratio method is undoubtedly a very familiar one, since the vast majority of tests in present statistical practice are based on this method. But for the sake of completeness we shall outline it briefly. Let the distribution law of the population  $K$  be of the form  $f(x^1, \dots, x^k; \theta_1, \dots, \theta_k)$  where the  $\theta$ 's may correspond to any point in a set  $\Omega$ , and let the hypothesis  $H$  to be tested be that the  $\theta$ 's actually belong to the subset  $\omega$  of  $\Omega$ . Form the likelihood function

$$P_N(x; \theta) = \prod_{\alpha=1}^N f(x_{\alpha}^1, \dots, x_{\alpha}^k; \theta_1, \dots, \theta_k)$$

i.e., the elementary probability law of a sample of  $N$  elements drawn independently from  $K$ . Denote by  $P_N^{\Omega}(x)$  the maximum of  $P_N$  for fixed  $x$  where the  $\theta$ 's are allowed to range over  $\Omega$ ; and denote by  $P_N^{\omega}(x)$  the corresponding maximum value when the  $\theta$ 's are restricted to  $\omega$ . The test criterion is then

$$\lambda = \frac{P_N^{\omega}(x)}{P_N^{\Omega}(x)}.$$

Evidently  $\lambda$  depends only on the observable quantities  $x'_\alpha$ , and has the range  $0 \leq \lambda \leq 1$ , with a definite probability law depending on that of the basic population  $K$ . In this method the critical region  $S$  is taken to be  $0 \leq \lambda \leq \lambda_c$ , where  $\lambda_c$  is so chosen that the probability  $P\{\lambda \leq \lambda_c\}$  is  $\epsilon$  when the parameters of  $K$  correspond to a point in  $\omega$ . (It may be noted here that in all the cases with which we shall have to deal the probability that  $\lambda$  lies in  $S$  when  $H$  is true is independent of the particular values of the  $\theta$ 's as long as they correspond to a point of  $\omega$ .) The reason for taking the critical region to be of the form  $0 \leq \lambda \leq \lambda_c$  and not, say,  $\lambda'_c \leq \lambda \leq \lambda''_c$  or  $\bar{\lambda}_c \leq \lambda \leq 1$  may become clearer when we examine the resulting tests for bias.

The recent work of Neyman and Pearson [10] has led them to lay considerable stress on the importance of unbiased tests. And though their attention has been directed mainly to the broader outlines of the theory of testing hypotheses, they have stimulated other writers to study particular tests of great practical importance. P. C. Tang [11] has obtained the general sampling distribution of  $1 - \lambda^{2/N}$  for what we shall call the regression problem with one dependent variate, and has given tables for  $P\{\lambda \leq \lambda_c\}$ —essentially proving the unbiased character of the test—which should be extremely useful. His article also contains an excellent discussion of the manner in which this test is related to the well known tests of linear hypotheses [7] and to the ordinary analysis of variance. P. L. Hsu [6] has shown that this same distribution is fundamental in the study of Hotelling's generalized  $T$  test [5] (a special but important case of what we shall call the general regression problem), and has proved that (locally) this test is not only unbiased but "most powerful" in a certain sense. On the other hand, it is not true that all likelihood ratio tests are unbiased [2]. Consequently, the knowledge that in a rather wide class of problems which arise in normal sampling theory the method of likelihood ratios furnishes tests which are either locally or completely unbiased would seem to be of some value, even when the exact sampling distribution of the criterion is too complicated to tabulate.

**2. The regression problem with one dependent variate.** Suppose that  $y$  is known to be normally distributed about a linear function of the fixed variables  $x^1, \dots, x^r$ , so that the family of populations under consideration is characterized by a distribution function of the form

$$(2.1) \quad f(y|x, b, \sigma^2) = (2\pi\sigma^2)^{-1} e^{-\frac{1}{2\sigma^2} \left( y - \sum_{i=1}^r b_i x^i \right)^2},$$

where the set of admissible values of  $\sigma^2$  and the  $b$ 's is

$$\Omega: 0 < \sigma^2 < \infty, \quad -\infty < b_i < \infty.$$

Let  $H$  be the hypothesis that the point  $(\sigma^2, b_1, \dots, b_r)$  lies in the subset of  $\Omega$  defined by

$$\omega: b_{q+1} = b_{q+2} = \dots = b_r = 0.$$

The likelihood ratio appropriate to testing the hypothesis  $H$  on the basis of  $N$  ( $N > r$ ) independent observations drawn from such a population is then

$$\lambda = \frac{\sum_{\alpha=1}^N \left( y_{\alpha} - \sum_{i=1}^r b_i x_{\alpha}^i \right)^2}{\sum_{\alpha=1}^N \left( y_{\alpha} - \sum_{k=1}^q b_k x_{\alpha}^k \right)^2} \Bigg\}^{1/N},$$

with the understanding that the values of the fixed variables  $x_{\alpha}^1, \dots, x_{\alpha}^r$  associated with the  $\alpha$ -th observation have been so chosen that the matrix  $\|a^{ij}\| = \sum_{\alpha=1}^N x_{\alpha}^i x_{\alpha}^j$  is positive definite. (The expression in the numerator is the minimum of  $\sum_{\alpha=1}^N \left( y_{\alpha} - \sum_{i=1}^r b_i x_{\alpha}^i \right)^2$  for variations of the  $b$ 's over  $\Omega$ , while the denominator contains the corresponding minimum for variations of the  $b$ 's over  $\omega$ ).

In order to show that the test is unbiased, we shall make use of the exact sampling distribution of the quantity

$$\xi = 1 - \lambda^{2/N},$$

first published by P. C. Tang [11]. Writing  $\|A_{\rho h}\|$  for the inverse of the matrix  $\|a^{\rho h}\|$  composed of the first  $q$  rows and columns of  $\|a^{ij}\|$ , let us put

$$G = \frac{1}{2\sigma^2} \sum_{k,l=q+1}^r \left( a^{kl} - \sum_{\rho,h=1}^q a^k A_{\rho h} a^{hl} \right) b_k b_l.$$

Since the critical region  $0 \leq \lambda \leq \lambda_*$  corresponds to the region  $1 - \lambda_*^{2/N} = \xi_* \leq \xi \leq 1$ , it can then be shown that the probability of rejecting  $H$  when the population parameters have specified values  $\sigma^2, b^1, \dots, b^r$  is expressed by the series

$$(2.2) \quad I(G, \xi_*) = e^{-G} \sum_{r=0}^{\infty} \frac{G^r}{r!} \int_{\xi_*}^1 \frac{\xi^{r(q-r)+r-1} (1-\xi)^{r(N-r)-1}}{B[\frac{1}{2}(r-q) + \nu, \frac{1}{2}(N-r)]} d\xi,$$

where

$$B(u, v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)} = \int_0^1 z^{u-1} (1-z)^{v-1} dz.$$

Now  $G$  is a positive definite quadratic form in the parameters  $b^{q+1}, \dots, b^r$ , so that it vanishes if and only if the hypothesis is true. And if  $0 < \epsilon < 1$ , then  $I(G, \xi_*)$  is a monotone increasing function of  $G$ . For by differentiating (2.2) we obtain

$$(2.3) \quad \frac{\partial}{\partial G} I(G, \xi_*) = e^{-G} \sum_{r=0}^{\infty} \frac{G^r}{r!} \int_{\xi_*}^1 \left\{ \frac{\xi^{r(q-r)+r} (1-\xi)^{r(N-r)-1}}{B[\frac{1}{2}(r-q) + \nu + 1, \frac{1}{2}(N-r)]} - \frac{\xi^{r(q-r)+r-1} (1-\xi)^{r(N-r)-1}}{B[\frac{1}{2}(r-q) + \nu, \frac{1}{2}(N-r)]} \right\} d\xi.$$

And from a property of incomplete Beta functions, which we shall demonstrate in the next section, it follows that each term in the series (2.3) is positive. Accordingly we have

**THEOREM I.** *The likelihood ratio test for the hypothesis that in a population of type (2.1) certain of the regression coefficients are zero, i.e., the hypothesis that  $y$  is independent of the fixed variables  $x^{t+1}, \dots, x^r$ , is completely unbiased.*

Wilks [15] has noted that the ordinary analysis of variance and covariance amounts essentially to testing hypotheses of this nature by means of the function

$$\zeta = \frac{1 - \lambda^{2/N}}{\lambda^{2/N}}.$$

Consequently such tests are also completely unbiased, since the region of rejection is then taken to be of the form  $\zeta \geq \zeta_*$ .

**3. An inequality relating to incomplete Beta functions.** Let us write

$$B(u, v; t) = \int_t^1 z^{u-1}(1-z)^{v-1} dz \quad (0 \leq t \leq 1).$$

Now,

$$\int_t^1 z^{u-1}(1-z)^v dz = \frac{z^u(1-z)^v}{u} \Big|_t^1 + \frac{v}{u} \int_t^1 z^u(1-z)^{v-1} dz.$$

The integrated term on the right is non-positive, so that

$$(3.1) \quad B(u, v+1; t) \leq \frac{v}{u} B(u+1, v; t)$$

in which the equality holds if and only if  $t = 0$  or  $t = 1$ . Again, since

$$z^u(1-z)^{v-1} + z^{u-1}(1-z)^v = z^{u-1}(1-z)^{v-1},$$

we have

$$(3.2) \quad B(u+1, v; t) + B(u, v+1; t) = B(u, v; t).$$

Combining these results, we find that

$$(3.3) \quad \frac{u+v}{u} B(u+1, v; t) \geq B(u, v; t)$$

with equality only when  $t = 0$  or  $t = 1$ . Hence we have

**LEMMA 1:** *If  $0 < t < 1$ , then*

$$\frac{B(u+1, v; t)}{B(u+1, v)} > \frac{B(u, v; t)}{B(u, v)}.$$

**4. The multiple correlation coefficient.** Suppose the distribution law of the underlying population is known to be of the form

$$(4.1) \quad f(x^1, \dots, x^t | x^{t+1}, \dots, x^m) = \frac{|B_{tt}|}{\pi^{1/2}} e^{-B_{tt}(x^t - a^t - C_{tt}^t)(x^t - a^t - C_{tt}^t)}.$$

The indices appearing in this expression take the values  $i, j = 1, \dots, t$  and  $p, q = t+1, \dots, m$ . The summation convention of repeated indices will be

used, for example,  $\sum_{p=i+1}^m C_p^i x^p$  will be denoted by  $C_p^i x^p$ . We shall also have occasion to use indices  $r, s$  with the range  $r, s = 1, \dots, m$ . The set of possible values of the  $a$ 's,  $B$ 's, and  $C$ 's is

$$\Omega: \|B_{ij}\| \text{ positive definite}; -\infty < a^i < \infty; -\infty < C_p^i < \infty.$$

We shall consider the  $\lambda$  test for the hypothesis  $H$  that  $x^1$  is independent of the remaining variables  $x^2, \dots, x^m$ , i.e., that the parameters belong to that subset of  $\Omega$  defined by

$$\omega: B_{1k} = 0, \quad (k = 2, \dots, t); \quad C_p^1 = 0.$$

Let us write  $v^{rs} = \sum_{\alpha=1}^N (x_\alpha^r - \bar{x}^r)(x_\alpha^s - \bar{x}^s)$ , and assume that the values of the fixed variables  $x_\alpha^p$  have been so selected that the matrix  $\|v^{pq}\|$  is positive definite. The likelihood ratio can then be expressed in the form

$$\lambda = \left( \frac{|v^{rs}|}{v^{11} \cdot \bar{v}_{11}} \right)^{\frac{1}{2}N} (1 - R^2)^{\frac{1}{2}N},$$

where  $\bar{v}_{11}$  is the complement of  $v^{11}$  in the determinant  $|v^{rs}|$ . If  $N \geq m + 1$ , the general sampling distribution of  $R^2$  (the multiple correlation coefficient between  $x^1$  and  $m - 1$  other variates), for this case in which  $x^2, \dots, x^t$  are subject to sampling variation and the remainder are fixed, is

$$(4.2) \quad F(R^2) d(R^2) = \frac{(1 - \rho^2)^{\frac{1}{2}(N-1)} e^{-\frac{1}{2}v^2} (1 - R^2)^{\frac{1}{2}(N-m)-1} (R^2)^{\frac{1}{2}(m-1)-1}}{\Gamma[\frac{1}{2}(N - m)]} \\ \times \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} \frac{\frac{1}{2}(\bar{y}^2)^\mu (1 - \rho^2)^\mu (\rho^2)^\nu (R^2)^{\mu+\nu} \Gamma^2[\frac{1}{2}(N - 1) + \mu + \nu]}{\mu! \nu! \Gamma[\frac{1}{2}(N - 1) + \mu] \Gamma[\frac{1}{2}(m - 1) + \mu + \nu]} d(R^2),$$

where

$$1 - \rho^2 = \frac{|B_{ij}|}{B_{11} \bar{B}^{11}}, \quad \frac{1}{2} \bar{y}^2 = \frac{v^{pq}}{B^{11}} C_p^1 C_q^1, \quad \|B^{ij}\| = \|B_{ij}\|^{-1}.$$

This distribution was first obtained by Wilks [13], although Fisher [3] had previously treated the two extreme cases in which (1) all independent variables are subject to sampling fluctuation, and (2) all independent variables are fixed.

To simplify the presentation, let us put  $\bar{\rho} = \rho^2$ ,  $\bar{y} = \frac{1}{2} \bar{y}^2$  and  $\bar{R} = R^2$ , and note that  $\bar{y} = 0$  if and only if  $C_p^1 = 0$  ( $p = t + 1, \dots, m$ ) while  $\bar{\rho} = 0$  if and only if  $B_{1k} = 0$  ( $k = 2, \dots, t$ ), so that  $\bar{y} = \bar{\rho} = 0$  means that the hypothesis  $H$  is true. On any alternative hypothesis, one or the other or both of these quantities will be positive. Let the region of rejection be taken to be

$$R_* \leq \bar{R} \leq 1,$$

which corresponds to

$$0 \leq \lambda \leq (1 - \bar{R}_*)^{\frac{1}{2}N}.$$

The probability of rejecting  $H$  is then

$$(4.4) \quad I(\bar{p}, \bar{y}, \bar{R}_s) = e^{-\bar{y}} \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} \frac{\bar{y}^{\mu}}{\mu!} (1 - \bar{p})^{\frac{1}{2}(N-1)+\mu} \frac{\bar{p}^{\nu} \Gamma[\frac{1}{2}(N-1) + \mu + \nu]}{\nu! \Gamma[\frac{1}{2}(N-1) + \mu]} \\ \times \int_{\bar{R}_s}^1 \frac{\bar{R}^{\frac{1}{2}(m-1)+\mu+\nu-1} (1 - \bar{R})^{\frac{1}{2}(N-m)-1}}{B[\frac{1}{2}(m-1) + \mu + \nu, \frac{1}{2}(N-m)]} d\bar{R}.$$

We shall show that  $I(\bar{p}, \bar{y}, \bar{R}_s)$  is a strictly monotone increasing function of  $\bar{p}$  for each  $\bar{y}$ , and that  $I(0, \bar{y}, \bar{R}_s)$  is a strictly monotone increasing function of  $\bar{y}$ .

First consider  $\frac{\partial I}{\partial \bar{p}}$ . We can write (4.4) in the form

$$I(\bar{p}, \bar{y}, \bar{R}_s) = e^{-\bar{y}} \sum_{\mu=0}^{\infty} \frac{\bar{y}^{\mu}}{\mu!} \frac{1}{\Gamma[\frac{1}{2}(N-1) + \mu]} \cdot \sum_{\nu=0}^{\infty} \frac{\bar{p}^{\nu}}{\nu!} (1 - \bar{p})^{\frac{1}{2}(N-1)+\mu} \varphi_{\mu, \nu},$$

where

$$\varphi_{\mu, \nu} = \Gamma[\frac{1}{2}(N-1) + \mu + \nu] \frac{B[\frac{1}{2}(m-1) + \mu + \nu, \frac{1}{2}(N-m); \bar{R}_s]}{B[\frac{1}{2}(m-1) + \mu + \nu, \frac{1}{2}(N-m)]}.$$

Then, formally,

$$\frac{\partial}{\partial \bar{p}} \left( \sum_{\nu=0}^{\infty} \frac{\bar{p}^{\nu}}{\nu!} (1 - \bar{p})^{\frac{1}{2}(N-1)+\mu} \varphi_{\mu, \nu} \right) \\ = \sum_{\nu=0}^{\infty} \frac{\nu \bar{p}^{\nu-1}}{\nu!} (1 - \bar{p})^{\frac{1}{2}(N-1)+\mu} - \sum_{\nu=0}^{\infty} \frac{\bar{p}^{\nu}}{\nu!} (1 - \bar{p})^{\frac{1}{2}(N-1)+\mu-1} [\frac{1}{2}(N-1) + \mu] \varphi_{\mu, \nu}.$$

Taking out the factor  $(1 - \bar{p})^{\frac{1}{2}(N-1)+\mu-1}$ , we have left

$$\sum_{\nu=0}^{\infty} \frac{\nu \bar{p}^{\nu-1}}{\nu!} \varphi_{\mu, \nu} - \sum_{\nu=0}^{\infty} \frac{\nu \bar{p}^{\nu}}{\nu!} \varphi_{\mu, \nu} - \sum_{\nu=0}^{\infty} \frac{\bar{p}^{\nu}}{\nu!} [\frac{1}{2}(N-1) + \mu] \varphi_{\mu, \nu} \\ = \sum_{\nu=0}^{\infty} \frac{\bar{p}^{\nu}}{\nu!} \{ \varphi_{\mu, \nu+1} - [\frac{1}{2}(N-1) + \mu + \nu] \varphi_{\mu, \nu} \}.$$

And the expression  $\varphi_{\mu, \nu+1} - [\frac{1}{2}(N-1) + \mu + \nu] \varphi_{\mu, \nu}$  is the same as

$$\Gamma[\frac{1}{2}(N-1) + \mu + \nu + 1] \left\{ \frac{B[\frac{1}{2}(m-1) + \mu + \nu + 1, \frac{1}{2}(N-m), \bar{R}_s]}{B[\frac{1}{2}(m-1) + \mu + \nu + 1, \frac{1}{2}(N-m)]} \right. \\ \left. - \frac{B[\frac{1}{2}(m-1) + \mu + \nu, \frac{1}{2}(N-m), \bar{R}_s]}{B[\frac{1}{2}(m-1) + \mu + \nu, \frac{1}{2}(N-m)]} \right\}$$

and is therefore positive, by Lemma 1. Consequently

$$\frac{\partial}{\partial \bar{p}} I(\bar{p}, \bar{y}, \bar{R}_s) \geq 0,$$

with equality holding only if  $\bar{p} = 1$ , or if the critical region is taken as the whole interval or the null set.

We have yet to investigate  $\frac{\partial}{\partial g} I(0, g, R_s)$ . In this case (4.4) becomes

$$(4.5) \quad I(0, g, R_s) = e^{-g} \sum_{\mu=0}^{\infty} \frac{g^{\mu}}{\mu!} \frac{B[\frac{1}{2}(m-1) + \mu, \frac{1}{2}(N-m), R_s]}{B[\frac{1}{2}(m-1) + \mu, \frac{1}{2}(N-m)]}.$$

(Note that this agrees with (2.2) if we make use of the relations  $r = m, q = 1$ , and  $B^{11} = 2\sigma^2$ .) We then obtain

$$\begin{aligned} \frac{\partial}{\partial g} I(0, g, R_s) = e^{-g} \sum_{\mu=0}^{\infty} \frac{g^{\mu}}{\mu!} & \left\{ \frac{B[\frac{1}{2}(m-1) + \mu + 1, \frac{1}{2}(N-m); R_s]}{B[\frac{1}{2}(m-1) + \mu + 1, \frac{1}{2}(N-m)]} \right. \\ & \left. - \frac{B[\frac{1}{2}(m-1) + \mu, \frac{1}{2}(N-m); R_s]}{B[\frac{1}{2}(m-1) + \mu, \frac{1}{2}(N-m)]} \right\} \end{aligned}$$

which the lemma shows to be positive when  $0 < R_s < 1$ .

This concludes the proof of

**THEOREM II.** *If the underlying population has a distribution law of the form (4.1), then the likelihood ratio test for the hypothesis that  $x^1$  is independent of  $x^2, \dots, x^m$ , where  $x^{t+1}, \dots, x^m$  are fixed and  $x^2, \dots, x^t$  are subject to sampling variation, is completely unbiased.*

**5. Mutual independence of several sets of random variables.<sup>1</sup>** Let the distribution law of the  $m$ -variate population be of the form

$$(5.1) \quad \frac{|B_{ij}|^{\frac{1}{2}}}{\pi^{\frac{1}{2}m}} e^{-B_{ij}(x^i - a^i)(x^j - a^j)}.$$

Here  $\Omega$  is the set  $\|B_{ij}\|$  positive definite;  $-\infty < a^i < \infty$ . Suppose we wish to test the hypothesis  $H_I$  that the variates  $\{x^1, \dots, x^{m_1}\}, \dots, \{x^{m_{p-1}+1}, \dots, x^{m_p}\}$  are mutually independent in sets [14], where  $0 = m_0 < m_1 < \dots < m_p = m$ . Then the  $\omega$  set is that defined by

$$\|B_{ij}\| = \|B_{i_1 j_1}\| + \dots + \|B_{i_p j_p}\| = \|B_1\| + \dots + \|B_p\|,$$

that is, we have  $B_{ij} = 0$  unless the indices  $i$  and  $j$  both relate to the same set of variates.

Associated with the population of random samples  $O_N$  ( $N \geq m + 1$ ) drawn from a universe characterized by (5.1), we have the distribution function

$$P(x; B, a) = \frac{|B_{ij}|^{\frac{1}{2}N}}{\pi^{\frac{1}{2}Nm}} e^{-\sum_{i=1}^N B_{ij}(x_a^i - a^i)(x_a^j - a^j)}.$$

The maximum of  $P$  with respect to variations of the parameters  $B_{ij}, a^i$  in  $\Omega$  is

$$P_{\Omega} = |v^{ij}|^{-\frac{1}{2}N} \left(\frac{N}{2\pi}\right)^{\frac{1}{2}Nm} e^{-\frac{1}{2}N},$$

<sup>1</sup> In this and in subsequent sections an index occurring both above and below indicates summation in accordance with the usual convention.

where

$$v^{ij} = \sum_{a=1}^N (x_a^i - \bar{x}^i)(x_a^j - \bar{x}^j).$$

And the maximum when the parameters are restricted to  $\omega$  is

$$P_\omega = [v_1 \dots v_p]^{-iN} \left( \frac{N}{2\pi} \right)^{iNm} e^{-iN},$$

where  $v_\mu$  stands for the determinant of the  $v$ 's connected with the  $\mu$ -th set of  $x$ 's. Thus the appropriate likelihood-ratio is given by

$$\lambda_I^{2/N} = \frac{|v^{ij}|}{v_1 \dots v_p}.$$

It is easy to see that the value of  $\lambda_I$  is unaltered if we replace  $x^i - a^i$  by  $x^i$ , so that we can express the probability that  $\lambda_I$  will lie between 0 and  $\lambda_*$  in the form

$$I(B, \lambda_*) = \frac{B^{iN}}{\pi^{iNm}} \int_{\lambda < \lambda_*} e^{-\sum_{a=1}^N B_{ij} x_a^i x_a^j} dx_1^1 \dots dx_N^m.$$

Furthermore,  $\lambda_I$  is invariant under the operation of replacing any  $x$  by a linear combination of  $x$ 's belonging to the same set. And since the assumption that  $\|B_{ij}\|$  is positive definite implies that the matrices  $\|B_{i_\mu j_\mu}\|$  have the same property, we can transform the  $x$ 's in each set among themselves by orthogonal transformations in such a way as to reduce each of the expressions

$$B_{i_\mu j_\mu} x^{i_\mu} x^{j_\mu}$$

to sums of squares. Thus we have

$$(5.2) \quad I(B, \lambda_*) = \frac{B^{*iN}}{\pi^{iNm}} \int_{\lambda < \lambda_*} e^{-\sum_{a=1}^N B_{i_\mu j_\mu}^* x_a^{i_\mu} x_a^{j_\mu}} dx_1^1 \dots dx_N^m = I(B^*, \lambda_*),$$

where

$$(5.3) \quad B_{i_\mu j_\mu}^* = \alpha_{i_\mu}^{h_\mu} B_{h_\mu k_\mu} \alpha_{j_\mu}^{k_\mu} \quad (h_\mu, i_\mu, j_\mu, k_\mu = m_{\mu-1} + 1, \dots, m_\mu),$$

$$(5.4) \quad B_{i_\sigma j_\tau}^* = 0 \quad i_\sigma \neq j_\tau,$$

and the subscripts on the indices indicate the sets of values over which they range; e.g.,  $i_2$  runs over the numbers corresponding to the columns of the matrix  $\|B_2\|$ . From (5.3) and (5.4) it is clear that  $\|B_{i_\mu j_\mu}^*\|$  reduces to a diagonal matrix when  $H$  is true.

In order to show that the test is locally unbiased, we may consider the derivatives

$$\left( \frac{\partial}{\partial B_{i_\mu j_\mu}^*} I(B^*, \lambda_*) \right)_0, \quad \left( \frac{\partial^2}{\partial B_{i_\mu j_\mu}^* \partial B_{h_\sigma k_\sigma}^*} I(B^*, \lambda_*) \right)_0, \quad (\mu \neq \nu, \sigma \neq \tau)$$



for the  $B^*$ 's are linear functions of the  $B$ 's; and the positive definiteness of one matrix of second partials implies that of the other. We have at once

$$\left( \frac{\partial B^*}{\partial B_{i,\mu}^*} \right)_0 = 0, \quad \left( \frac{\partial^2 B^*}{\partial B_{i,\mu}^* \partial B_{h,\nu}^*} \right)_0 = 0, \quad (\mu \neq \nu, \sigma \neq \tau)$$

unless the second derivative is taken twice with respect to the same  $B^*$ . Thus

$$\left( \frac{\partial I(B^*, \lambda_s)}{\partial B_{i,\mu}^*} \right)_0 = -2 \frac{B_0^{*1N}}{\pi^{1Nm}} \int_{\lambda < \lambda_s} \sum_{\alpha=1}^N x_{\alpha}^{i_{\mu}} x_{\alpha}^{j_{\mu}} e^{-\sum_{\alpha=1}^N B_{i,j}^* x_{\alpha}^{i_{\mu}} x_{\alpha}^{j_{\mu}}} dx,$$

where the  $B_0^*$  indicates that the  $B$ 's have the diagonal form associated with  $H$ . And since whenever the point  $x_1^1, \dots; x_1^i, \dots, x_N^i; \dots, x_N^m$  is in the region  $\lambda \leq \lambda_s$ , so also is the point  $x_1^1, \dots; -x_1^i, \dots, -x_N^i; \dots, x_N^m$  it follows that

$$\frac{\partial}{\partial B_{i,\mu}^*} I(B_0^*, \lambda_s) = 0, \quad (\mu \neq \nu).$$

Similar considerations show that the non-repeated second derivatives

$$\frac{\partial^2}{\partial B_{i,\mu}^* \partial B_{h,\nu}^*} I(B_0^*, \lambda_s) = 4 \frac{B_0^{*1N}}{\pi^{1Nm}} \int_{\lambda < \lambda_s} \left( \sum_{\alpha=1}^N x_{\alpha}^{i_{\mu}} x_{\alpha}^{j_{\mu}} \right) \left( \sum_{\beta=1}^N x_{\beta}^{h_{\nu}} x_{\beta}^{k_{\nu}} \right) e^{-\sum_{\alpha=1}^N B_{i,j}^* x_{\alpha}^{i_{\mu}} x_{\alpha}^{j_{\mu}}} dx$$

must vanish.

Finally, we must show that the repeated second derivatives are positive when evaluated at a point in  $\omega$ , except of course in the trivial cases  $\lambda_s = 0, \lambda_s = 1$ , when they must be zero. In order to do this, we shall make use of the fact that the  $v$ 's which go to make up  $\lambda$  have the Wishart distribution [17]

$$(5.5) \quad \frac{B^{1(N-1)}}{\pi^{\frac{1}{2}(m(m-1))} \prod_{i=1}^m \Gamma[\frac{1}{2}(N-i)]} \cdot v^{1(N-m)-1} e^{-B_{i,j} v^{ij}} dv^{11} \dots dv^{mm}.$$

(Because of the relation  $v^{ij} = v^{ji}$ , only  $\frac{1}{2}m(m+1)$  of the  $v$ 's appear as differentials). It will be useful to have the notation

$$G(B, N-1, m) = \frac{B^{1(N-1)}}{\pi^{\frac{1}{2}(m(m-1))} \prod_{i=1}^m \Gamma[\frac{1}{2}(N-i)]}$$

$$V(B, N-1, m) = v^{1(N-m)-1} e^{-B_{i,j} v^{ij}}.$$

With the aid of (5.5) we shall now compute the moments

$$E[(\lambda^{2/N})^h], \quad h = 0, 1, \dots,$$

for the case in which the matrix  $\|B_{ij}\|$  has the form

$$(5.6) \quad \left\| \begin{array}{cccccc} B_{11} & \dots & B_{1m_1} & 0 & \dots & 0B_{1m} \\ \vdots & & \vdots & \vdots & & \vdots \\ B_{m_1 1} & \dots & B_{m_1 m_1} & 0 & \dots & 0 \\ 0 & \dots & 0 & & & \\ \vdots & & \vdots & & & \\ 0 & & & & \|B\| & \\ B_{m1} 0 & \dots & 0 & & & \end{array} \right\|$$

where  $\|B\|$  stands for  $\|B_2\| + \dots + \|B_p\|$ , and all other  $B$ 's, except those indicated, are zero. Let us designate by  $(\bar{v})$  the set of  $v^{ij}$  which correspond to the rows and columns of  $\bar{B}$ , and by  $(v - \bar{v})$  the remaining  $v$ 's. We then remark that the result of integrating (5.5) with respect to the  $v$ 's in  $(v - \bar{v})$  is to reduce it to the corresponding distribution for the variables in the set  $\bar{v}$ , thus:

$$(5.7) \quad G(B, N - 1, m) \int V(B, N - 1, m) d(v - \bar{v}) \\ = G(\bar{B}, N - 1, m - m_1) V(\bar{B}, N - 1, m - m_1),$$

where  $\|\bar{B}_{kl}\|$  is the inverse of the matrix obtained by inverting  $\|B_{ij}\|$ , and striking out the first  $m_1$  rows and columns, that is

$$\bar{B}^{kl} = B^{kl}, \quad (k, l = m_1 + 1, \dots, m).$$

Then,

$$G(B, N - 1, m) \int \frac{v^k}{v_1^k \dots v_p^k} V(B, N - 1, m) d(v - \bar{v})$$

can be written as

$$(5.8) \quad \frac{G(B, N - 1, m)}{G(B, N - 1 + 2h, m)} \cdot G(B, N - 1 + 2h, m) \int v_1^{-k} \dots v_p^{-k} \\ \times V(B, N - 1 + 2h, m) d(v - \bar{v}) \\ = \frac{G(B, N - 1, m)}{G(B, N - 1 + 2h, m)} G(\bar{B}, N - 1 + 2h, m - m_1) \\ \times v_1^{-k} \dots v_p^{-k} V(\bar{B}, N - 1 + 2h, m - m_1).$$

It can be seen from (5.6) that

$$\|\bar{B}\| = \|B_2\| + \dots + \|B_{p-1}\| + \|\bar{B}_p\|$$

since of all the rows and columns of  $\|B_{ij}\|$  which are involved in  $\|\bar{B}\|$  it is only the last in which a non zero element appears outside of the blocks  $\|B_2\|$ ,  $\dots$ ,  $\|B_p\|$ . Consequently, the  $v$ 's corresponding to the determinants  $v_1, \dots$ ,

$v_p$  are independently distributed, so that if in (5.8) we integrate out all the remaining  $v$ 's but these, we shall be left with a product of factors

$$\frac{G(B, N-1, m)}{G(B, N-1+2h, m)} \cdot \prod_{i=2}^{p-1} \frac{G(B_i, N-1+2h, k_i)}{G(B_i, N-1, k_i)} \\ \times G(B_i, N-1, k_i) v_i^{-h} V(B_i, N-1+2h, k_i) \\ \times \frac{G(\tilde{B}_p, N-1+2h, k_p)}{G(\tilde{B}_p, N-1, k_p)} \cdot G(\tilde{B}_p, N-1, k_p) v_p^{-h} V(\tilde{B}_p, N-1+2h, k_p),$$

where  $k_\mu$  stands for the order of  $\|B_\mu\|$ . And this, when integrated with respect to the  $v$ 's in  $v_2, \dots, v_p$ , yields

$$\frac{G(B, N-1, m)}{G(B, N-1+2h, m)} \cdot \prod_{i=2}^{p-1} \frac{G(B_i, N-1+2h, k_i)}{G(B_i, N-1, k_i)} \times \frac{G(\tilde{B}_p, N-1+2h, k_p)}{G(\tilde{B}_p, N-1, k_p)},$$

which, because of the definition of the  $G$ 's, reduces to

$$\prod_{i=1}^m \frac{\Gamma[\frac{1}{2}(N-i)+h]}{\Gamma[\frac{1}{2}(N-i)]} \cdot \prod_{i=2}^p \prod_{i=1}^{k_i} \frac{\Gamma[\frac{1}{2}(N-i)]}{\Gamma[\frac{1}{2}(N-i)+h]} \times B^{-h} B_2^h \dots B_{p-1}^h \tilde{B}_p^h.$$

Denoting the product of ratios of  $\Gamma$ 's by  $K_h$ , and recalling the form of  $\|B_{ij}\|$ , we therefore have

$$(5.9) \quad E \left[ \frac{v^h}{v_2^h \dots v_p^h} \right] = K_h \tilde{B}_p^h B'^{-h}$$

with

$$\begin{array}{ccccccc} B_{11} & \dots & B_{1m_1} & 0 & \dots & 0 & B_{1m} \\ & & & & & & 0 \\ & & & & & & 0 \\ |B'| & B_{m_1 1} & \dots & B_{m_1 m_1} & 0 & \dots & 0 \\ & 0 & \dots & 0 & & & \\ & \vdots & & \vdots & & & \\ & 0 & & \vdots & & & \|B_p\| \\ & B_{m1} 0 & \dots & 0 & & & \end{array}$$

But it is not difficult to see that under the condition (5.6), the matrix  $\|\tilde{B}_p\|$  is also the inverse of the matrix obtained by striking out the first  $m$  rows and columns in the inverse of  $\|B'\|$ . Making use of this relation, we can apply the Jacobi theorem to (5.9), and put that expression in the form

$$E \left[ \frac{v^h}{v_2^h \dots v_p^h} \right] = K_h B_1^{-h},$$

where  $\|B_1\|$  is the matrix in the upper left hand corner of  $\|B'\|$ , namely  $\|B_{i_1 i_1}\|$ .

Let the subscript  $\beta$  on a  $B$  stand for the result of replacing  $B_{i_1 j_1}$  by  $B_{i_1 j_1} + \beta_{i_1 j_1}$ . For sufficiently small values of the  $\beta$ 's the matrix  $\|B_{i_1 j_1}\|$  will still be positive definite, so that we shall have

$$\frac{B_{\beta}^{1(N-1)}}{\pi^{1(m(m-1))} \prod_{i=1}^m \Gamma[\frac{1}{2}(N-i)]} \int \frac{v^{\frac{1}{2}}}{v_1^{\frac{1}{2}} \dots v_p^{\frac{1}{2}}} v^{1(N-m)-1} e^{-B_{i_1 j_1} v^{i_1 j_1}} dv = K_{\beta} B_{1\beta}^{-1},$$

which we can put in the form

$$(5.10) \quad K' \int \frac{v^{\frac{1}{2}}}{v_1^{\frac{1}{2}} \dots v_p^{\frac{1}{2}}} v^{1(N-m)-1} e^{-B_{i_1 j_1} v^{i_1 j_1}} dv = \frac{K_{\beta}}{B_{1\beta}^1 B_{\beta}^{1(N-1)}}.$$

Wilks [13] has shown how to generate moments of determinants by the device of replacing  $\beta_{i_1 j_1}$  by  $\beta_{i_1 j_1} + \xi_{i_1} \xi_{j_1}$ , and integrating with respect to the  $\xi$ 's from  $-\infty$  to  $\infty$ . Applying this process  $2h$  times to the left hand side of (5.10) gives

$$\pi^{1h} K' \int \left( \frac{v}{v_1 \dots v_p} \right)^h V(B_{\beta}, N-1, m) dv,$$

which when multiplied by  $\pi^{-1h} B^{1(N-1)}$  yields

$$E[(\lambda^{2/N})^h]$$

when the  $\beta$ 's are set equal to zero.

To obtain the value of this expression, we may perform the same operations on the right hand side of (5.10). But before so doing, we shall put  $B_{\beta}$  in a more convenient form. We have

$$B_{\beta} = B_{1\beta} \cdot B - B_{1m}^2 \cdot B B^{mm} \cdot B_{1\beta}'^{11},$$

where  $B^{mm}$  is the inverse element of  $B_{mm}$  in  $\|B\|$ , and  $B_{1\beta}'^{11}$  is the cofactor of  $B_{1\beta}$  in  $B_{1\beta}$ , the result being obtained by expanding  $B_{\beta}$  according to minors of the first row and first column. Similarly,

$$(5.11) \quad B = B_1 \cdot B - B_{1m}^2 \cdot B B^{mm} \cdot B_1'^{11}.$$

From (5.11) we have

$$\frac{B}{B_1 \dots B_p} = 1 - B_{1m}^2 B^{mm} \cdot \frac{B_1'^{11}}{B_1},$$

so that if we put  $B \cdot B_1^{-1} \dots B_p^{-1} = \Lambda$ , we find that

$$\begin{aligned} B_{\beta} &= B_{1\beta} \cdot B \left\{ 1 - B_{1m}^2 B^{mm} \cdot \frac{B_1'^{11}}{B_1} \cdot \frac{B_1}{B_1'^{11}} \cdot \frac{B_{1\beta}'^{11}}{B_{1\beta}} \right\} \\ &= B_{1\beta} B \left\{ 1 - \frac{B_1}{B_1'^{11}} (1 - \Lambda) \frac{B_{1\beta}'^{11}}{B_{1\beta}} \right\}. \end{aligned}$$

Thus the result of multiplying (5.10) through by  $B^{i(N-1)}$  (where no  $\beta$ 's are substituted in this determinant) can be put in the form

$$(5.12) \quad \left( \frac{B}{B_s \dots B_p} \right)^{i(N-1)} \left\{ 1 - \frac{B_1}{B_1^{i11}} (1 - \Lambda) \frac{B_1^{i11}}{B_1} \right\}^{-i(N-1)} B_{1\beta}^i.$$

Expanding the expression in curled brackets, we get

$$\Lambda^{i(N-1)} B_1^{i(N-1)} \sum_{\nu=0}^{\infty} \frac{\Gamma[\frac{1}{2}(N-1) + \nu]}{\nu! \Gamma[\frac{1}{2}(N-1)]} B_1^{\nu} \left( \frac{B_1^{i11}}{B_1} \right)^{\nu} B_{1\beta}^{-[i(N-1)+h+\nu]} (1 - \Lambda)^{\nu}.$$

If we let  $B_{1\beta t}$  stand for the result of replacing  $B_{11}$  by  $B_{11} - t$  in  $B_{1\beta}$ , we can write this as

$$(5.13) \quad \Lambda^{i(N-1)} \sum_{\nu=0}^{\infty} \frac{\Gamma[\frac{1}{2}(N-1) + \nu]}{\nu! \Gamma[\frac{1}{2}(N-1)]} (1 - \Lambda)^{\nu} (B_1^{i11})^{-\nu} B_1^{i(N-1)+\nu} \\ \times \frac{\Gamma[\frac{1}{2}(N-1) + h]}{\Gamma[\frac{1}{2}(N-1) + h + \nu]} \frac{\partial^{\nu}}{\partial t^{\nu}} B_{1\beta t}^{-[i(N-1)+h]}$$

the derivatives being evaluated at  $t = 0$ .

Now Wilks' results show that the operation of introducing  $\beta_{i_1 j_1} + \xi_{i_1} \xi_{j_1}$  into  $B_{1\beta t}$  to replace  $\beta_{i_1 j_1}$  and integrating with respect to the  $\xi$ 's, when repeated  $2h$  times on  $B_{1\beta t}^{-[i(N-1)+h]}$ , produces

$$\pi^{m_1 h} B_{1t}^{-i(N-1)} \prod_{i=1}^{m_1} \frac{\Gamma[\frac{1}{2}(N-i)]}{\Gamma[\frac{1}{2}(N-i) + h]}$$

when the  $\beta$ 's are finally set equal to zero. Reversing the order of summation, differentiation and integration in (5.13), we thus obtain

$$(5.14) \quad \pi^{m_1 h} \prod_{i=1}^{m_1} \frac{\Gamma[\frac{1}{2}(N-i)]}{\Gamma[\frac{1}{2}(N-i) + h]} \Lambda^{i(N-1)} \sum_{\nu=0}^{\infty} \frac{\Gamma[\frac{1}{2}(N-1) + \nu]}{\nu! \Gamma[\frac{1}{2}(N-1)]} \\ \times (1 - \Lambda)^{\nu} (B_1^{i11})^{-\nu} B_1^{i(N-1)+\nu} \frac{\Gamma[\frac{1}{2}(N-1) + h]}{\Gamma[\frac{1}{2}(N-1) + h + \nu]} \left( \frac{\partial^{\nu}}{\partial t^{\nu}} B_{1t}^{-i(N-1)} \right)_0.$$

Now

$$\left( \frac{\partial^{\nu}}{\partial t^{\nu}} B_{1t}^{-i(N-1)} \right)_0 = \frac{\Gamma[\frac{1}{2}(N-1) + \nu]}{\Gamma[\frac{1}{2}(N-1)]} \cdot (B_1^{i11})^{\nu} B_1^{-[i(N-1)+\nu]},$$

so that (5.14) becomes

$$\pi^{m_1 h} \prod_{i=1}^{m_1} \frac{\Gamma[\frac{1}{2}(N-i)]}{\Gamma[\frac{1}{2}(N-i) + h]} \Lambda^{i(N-1)} \sum_{\nu=0}^{\infty} \frac{\Gamma[\frac{1}{2}(N-1)]}{\nu! \Gamma[\frac{1}{2}(N-1)]} \\ \times (1 - \Lambda)^{\nu} \frac{\Gamma[\frac{1}{2}(N-1) + h]}{\Gamma[\frac{1}{2}(N-1) + h + \nu]} \cdot \frac{\Gamma[\frac{1}{2}(N-1) + \nu]}{\Gamma[\frac{1}{2}(N-1)]}.$$

From this it appears that the  $h$ -th moment of  $\lambda_i^{3/N}$  is given by

$$\begin{aligned}
 E[(\lambda^{3/N})^h] &= \prod_{i=1}^m \frac{\Gamma[\frac{1}{2}(N-i) + h]}{\Gamma[\frac{1}{2}(N-i)]} \cdot \prod_{i=1}^p \prod_{i=1}^{k_i} \frac{\Gamma[\frac{1}{2}(N-i)]}{\Gamma[\frac{1}{2}(N-i) + h]} \\
 (5.15) \quad &\times \Lambda^{k(N-1)} \sum_{v=0}^{\infty} (1-\Lambda)^v \frac{\Gamma[\frac{1}{2}(N-1) + v]}{v! \Gamma[\frac{1}{2}(N-1)]} \\
 &\times \frac{\Gamma[\frac{1}{2}(N-1) + v]}{\Gamma[\frac{1}{2}(N-1) + h + v]} \cdot \frac{\Gamma[\frac{1}{2}(N-1) + h]}{\Gamma[\frac{1}{2}(N-1)]}.
 \end{aligned}$$

A considerable amount of cancellation will take place in (5.15), for  $m$  is greater than any  $k_i$ . Suppose the largest  $k_i$  is  $k_{i'}$ . Then we can cancel its product into the first one, with the assurance that there will be at least one factor

$$(5.16) \quad \frac{\Gamma[\frac{1}{2}(N-1)]}{\Gamma[\frac{1}{2}(N-1) + h]}$$

to cancel the corresponding factor under the summation sign. Hence we have

$$\begin{aligned}
 E[(\lambda^{3/N})^h] &= \prod_{i=k_{i'}+1}^m \frac{\Gamma[\frac{1}{2}(N-i) + h]}{\Gamma[\frac{1}{2}(N-i)]} \cdot \prod_{i=1}^p \prod_{i=1}^{k_i} \frac{\Gamma[\frac{1}{2}(N-i)]}{\Gamma[\frac{1}{2}(N-i) + h]} \\
 (5.17) \quad &\times \Lambda^{k(N-1)} \sum_{v=0}^{\infty} (1-\Lambda)^v \frac{\Gamma[\frac{1}{2}(N-1) + v]}{v! \Gamma[\frac{1}{2}(N-1)]} \cdot \frac{\Gamma[\frac{1}{2}(N-1) + v]}{\Gamma[\frac{1}{2}(N-1) + h + v]},
 \end{aligned}$$

where  $\Pi'$  indicates that  $i'$  has been omitted, and  $\Pi''$  indicates that one factor (5.16) has been cancelled. Then we can take out the factor  $i = m$  in the first product, putting it under the summation sign, where, together with the final factor in each term of the sum, it gives rise to the combination

$$\frac{\Gamma[\frac{1}{2}(N-1) + v]}{\Gamma[\frac{1}{2}(N-m)] \Gamma[\frac{1}{2}(m-1) + v]} \cdot \frac{\Gamma[\frac{1}{2}(N-m) + h] \Gamma[\frac{1}{2}(m-1) + v]}{\Gamma[\frac{1}{2}(N-1) + h + v]}.$$

After making this reduction, we obtain

$$\begin{aligned}
 E[(\lambda^{3/N})^h] &= \prod_{i=k_{i'}+1}^{m-1} \frac{\Gamma[\frac{1}{2}(N-i) + h]}{\Gamma[\frac{1}{2}(N-i)]} \cdot \prod_{i=2}^p \prod_{i=1}^{k_i} \frac{\Gamma[\frac{1}{2}(N-i)]}{\Gamma[\frac{1}{2}(N-i) + h]} \\
 (5.18) \quad &\times \Lambda^{k(N-1)} \sum_{v=0}^{\infty} (1-\Lambda)^v \frac{\Gamma[\frac{1}{2}(N-1) + v]}{v! \Gamma[\frac{1}{2}(N-1)]} \frac{B[\frac{1}{2}(N-m) + h, \frac{1}{2}(m-1) + v]}{B[\frac{1}{2}(N-m), \frac{1}{2}(m-1) + v]}.
 \end{aligned}$$

The products of ratios in the first part of (5.18) are of the type discussed by Wilks in connection with integral equations of type  $B$  [12]. It follows from his results that  $\lambda_i^{3/N}$  is distributed like the product

$$z \cdot \theta_1 \dots \theta_{m'}, \quad (m' = m - k_{i'} - 1),$$

where  $z$  and the  $\theta$ 's are independently distributed, with the distribution of the  $\theta$ 's given by

$$f(\theta_1, \dots, \theta_{m'}) = \prod_{i=1}^{m'} \frac{\Gamma(c_i)}{\Gamma(b_i) \Gamma(c_i - b_i)} \cdot \theta_i^{b_i-1} (1 - \theta_i)^{c_i-b_i-1},$$

where the  $b_i$  and  $c_i$  are constants which depend on  $N$ ,  $m$ , and the sizes of the blocks, but not on  $\Lambda$ , and the distribution of  $z$  is given by

$$F(z) = \Lambda^{i(N-1)} \sum_{v=0}^{\infty} (1-\Lambda)^v \frac{\Gamma[\frac{1}{2}(N-1) + v]}{v! \Gamma[\frac{1}{2}(N-1)]} \cdot \frac{z^{i(N-m)-1} (1-z)^{i(m-1)+v-1}}{B[\frac{1}{2}(N-m), \frac{1}{2}(m-1) + v]}.$$

Consequently, the probability that  $\lambda$  lies between zero and  $\lambda_c$  is

$$J(\Lambda, \lambda_c) = \Lambda^{i(N-1)} \int_S \sum_{v=0}^{\infty} (1-\Lambda)^v \frac{\Gamma[\frac{1}{2}(N-1) + v]}{v! \Gamma[\frac{1}{2}(N-1)]} \\ \times f(\theta) \frac{z^{i(N-m)-1} (1-z)^{i(m-1)+v-1}}{B[\frac{1}{2}(N-m), \frac{1}{2}(m-1) + v]} dz d\theta,$$

where the integral is to be extended over the region

$$S: 0 \leq z \cdot \theta_1 \dots \theta_m < \lambda_c^{2/N}, \quad 0 \leq \theta_i \leq 1, \quad 0 \leq z \leq 1.$$

Let us integrate first with respect to  $z$  and then with respect to the  $\theta$ 's; we have

$$(5.19) \quad J(\Lambda, \lambda_c) = \int_{S_\theta} \Lambda^{i(N-1)} \sum_{v=0}^{\infty} (1-\Lambda)^v \frac{\Gamma[\frac{1}{2}(N-1) + v]}{v! \Gamma[\frac{1}{2}(N-1)]} \\ \times \frac{B'[\frac{1}{2}(N-m), \frac{1}{2}(m-1) + v; \varphi]}{B[\frac{1}{2}(N-m), \frac{1}{2}(m-1) + v]} f(\theta) d\theta,$$

where  $S_\theta$  is the set  $\Pi\theta_i < \lambda_c^{2/N}$ ,  $0 \leq \theta_i \leq 1$ , and

$$(5.20) \quad B'(u, v, \varphi) = \int_0^\varphi z^{u-1} (1-z)^{v-1} dz \\ = \int_{1-\varphi}^1 z^{v-1} (1-z)^{u-1} dz = B(v, u, 1-\varphi),$$

$\varphi(\theta)$  being the upper limit for  $z$  for fixed  $\theta$ . It is clear that the subset of  $s_\theta$  for which  $\varphi(\theta) < 1$  will not be of measure zero in the  $\theta$ -space, since we assume that  $0 < \lambda_c < 1$ .

The relation between (5.19) and the corresponding expression for the multiple correlation coefficient without fixed variates—the case  $\bar{y} = 0$  in (4.4)—may be clearer if we put

$$(5.21) \quad \bar{\rho} = 1 - \Lambda = B_{1m}^2 B^{mm} B_1^{11},$$

where  $B^{mm}$  is the inverse of  $B_{mm}$  in  $\|B\|$ , and  $B_1^{11}$  is the inverse of  $B_{11}$  in  $\|B_1\|$ . Then the required probability of rejection when  $\bar{\rho}$  has any fixed value is

$$I(\bar{\rho}, 1 - \lambda_c^{2/N}) = \int_{S_\theta} \sum_{v=0}^{\infty} \frac{\bar{\rho}^v}{v!} (1-\bar{\rho})^{i(N-1)} \frac{\Gamma[\frac{1}{2}(N-1) + v]}{\Gamma[\frac{1}{2}(N-1)]} \\ \times \frac{B[\frac{1}{2}(m-1) + v, \frac{1}{2}(N-m), 1-\varphi]}{B[\frac{1}{2}(m-1) + v, \frac{1}{2}(N-m)]} f(\theta) d\theta,$$

where we have used the relation (5.20) between the incomplete Beta functions. Differentiating with respect to  $\bar{\rho}$  before performing the integration with respect

to the  $\theta$ 's, we find by a computation similar to that in section 4 that each term in the series is positive except where  $\varphi(\theta) = 1$ ; so that we have

$$\frac{\partial I}{\partial \bar{\rho}}(\bar{\rho}, 1 - \lambda_e^{2/N}) > 0 \quad (\lambda_e \neq 1, 0).$$

And by (5.21), we then have

$$\frac{\partial^2 I}{\partial B_{im}^2} > 0.$$

Since the argument is clearly independent of which  $B_{i\mu}$ , ( $\mu \neq \nu$ ) we take, it follows that the test is locally unbiased. We have therefore proved:

**THEOREM III.** *If  $x^1, \dots, x^m$  have the joint normal distribution (5.1), then the likelihood ratio test for the hypothesis that the  $x$ 's are independent in sets is locally unbiased.*

In certain types of statistical material it may be important to consider, not the independence of the  $x$ 's themselves, but of their deviations from regression functions. For example, in the case of several related time series, it may be desirable to eliminate the trend of each  $x^i$  by means of, say, a second degree polynomial in  $t$ . Consider then in general a population whose distribution function is of the form

$$\frac{B^1}{\pi^{1/2m}} e^{-B_{ij}(x^i - C_{\mu}^i x^{\mu})(x^j - C_{\nu}^j x^{\nu})} \quad (\mu, \nu = m+1, \dots, m+q)$$

with unknown  $B_{ij}$  and  $C_{\mu}^i$ . The likelihood ratio for testing the hypothesis  $H_1$  that the sets of deviations

$$x^1 - C_{\mu}^1 x^{\mu}, \dots, x^{m_1} - C_{\mu}^{m_1} x^{\mu}; \dots; x^{m_{p-1}+1} - C_{\mu}^{m_{p-1}+1} x^{\mu}, \dots, x^m - C_{\mu}^m x^{\mu}$$

are independent is

$$\lambda_1 = \left\{ \frac{|d^{ij}|}{d_1 \dots d_p} \right\}^{1/2N}$$

where

$$d^{ij} = \Sigma (x_{\alpha}^i - \hat{C}_{\mu}^i x_{\alpha}^{\mu})(x_{\alpha}^j - \hat{C}_{\nu}^j x_{\alpha}^{\nu})$$

and  $\hat{C}_{\mu}^i$  is the usual least squares estimate of  $C_{\mu}^i$ , given by

$$\hat{C}_{\mu}^i a^{\mu\nu} = a^{i\nu}$$

with

$$a^{rs} = \Sigma x_{\alpha}^r x_{\alpha}^s \quad (r, s = 1, \dots, m+q).$$

An examination of the characteristic function of the  $d^{ij}$  shows that their distribution law is the same as that of the  $v^{ij}$  of the preceding discussion, except for the fact that  $N-1$  is replaced by  $N-q$ . Consequently the above results on freedom from bias, and also those of the next section, apply equally well to the  $\lambda_1$  test for the independence of deviations from regression functions.



6. On the moments of  $\lambda_I^{3/N}$ . Although we have succeeded in proving the unbiased nature of the preceding test only in the local sense, we can show that the moments of the criterion  $\lambda_I^{3/N}$  have a property which seems very closely related to that of furnishing a completely unbiased test. For it can be shown that each of the quantities

$$E[(\lambda_I^{3/N})^h] \quad h = \frac{1}{2}, 1, 1\frac{1}{2}, \dots$$

is greater, when  $H_I$  is true than when any alternative  $H'$  holds. It will perhaps be sufficient to prove this statement in detail for the case where  $h = 1$  and where  $H_I$  is the hypothesis that the matrix  $\|B_{ij}\|$  has the form  $\|\tilde{B}_0\| + \|B_{i_1j_1}\|$ :

$$\begin{array}{ccc} B_{11} & B_{12} & \\ & 0 & 0 \\ B_{21} & B_{22} & \\ & B_{33} & B_{34} \\ 0 & & 0 \\ & B_{43} & B_{44} \\ 0 & 0 & \|B_{i_1j_1}\| \end{array}$$

in the notation of the preceding section we then have

$$i_1, j_1 = 1, 2; \quad i_2, j_2 = 3, 4; \quad i_3, j_3 = 5, \dots, m.$$

Even when  $H$  is not true we find that

$$(6.1) \quad E[v^{ij} | v^{i_1j_1}] = \frac{G(B, N-1, m)}{G(\tilde{B}, N-1+2h, m)} \cdot \frac{G(\tilde{B}, N-1+2h, m-4)}{G(\tilde{B}, N-1, m-4)},$$

where  $\tilde{B}^{i_1j_1} = B^{i_1j_1}$ . Using the definition of the  $G$ 's in section 5 and the Jacobi theorem, we can write (6.1) in the form

$$E[v^{ij} | v^{i_1j_1}] = K_h \tilde{B}^{-h}$$

where  $\tilde{B}$  is the determinant of the matrix composed of the first four rows and columns of  $\|B_{ij}\|$ . In the general case we therefore have

$$\|\tilde{B}\| = \begin{vmatrix} B_{11} & B_{12} & B_{13} & B_{14} \\ B_{21} & B_{22} & B_{23} & B_{24} \\ B_{31} & B_{32} & B_{33} & B_{34} \\ B_{41} & B_{42} & B_{43} & B_{44} \end{vmatrix}$$

Thus if we set  $h = 1$ , and replace  $B_{i_1j_1}$  and  $B_{i_2j_2}$  by  $B_{i_1j_1} + \xi_{i_1}^{(1)}\xi_{j_1}^{(1)} + \xi_{i_1}^{(2)}\xi_{j_1}^{(2)}$  and  $B_{i_3j_3} + \xi_{i_3}^{(3)}\xi_{j_3}^{(3)} + \xi_{i_3}^{(4)}\xi_{j_3}^{(4)}$  respectively, indicating this replacement by a prime, we obtain

$$(6.2) \quad E[(\lambda_I^{3/N})^1] = K_1 \int B^{1(N-1)} B'^{1(N-1)} \tilde{B}'^{-1} d\xi.$$

Treating  $B'$  as a bordered determinant, we can reduce it to

$$\begin{aligned} B' &= B_{(12)}(1 + B_{(12)}^{i_1 i_2} \xi_{i_1}^{(4)} \xi_{i_2}^{(4)}) \\ &= B_{(12)}(1 + B_{(12)}^{i_1 i_2} \xi_{i_1}^{(3)} \xi_{i_2}^{(3)})(1 + B_{(12)}^{i_1 i_2} \xi_{i_1}^{(4)} \xi_{i_2}^{(4)}) \\ &= B_{(1)}(1 + B_{(1)}^{i_1 i_1} \xi_{i_1}^{(2)} \xi_{i_1}^{(2)})(1 + B_{(12)}^{i_1 i_2} \xi_{i_1}^{(3)} \xi_{i_2}^{(3)})(1 + B_{(12)}^{i_1 i_2} \xi_{i_1}^{(4)} \xi_{i_2}^{(4)}) \\ &= B(1 + B_{(1)}^{i_1 i_1} \xi_{i_1}^{(1)} \xi_{i_1}^{(1)})(1 + B_{(1)}^{i_1 i_1} \xi_{i_1}^{(2)} \xi_{i_1}^{(2)})(1 + B_{(12)}^{i_1 i_2} \xi_{i_1}^{(3)} \xi_{i_2}^{(3)})(1 + B_{(12)}^{i_1 i_2} \xi_{i_1}^{(4)} \xi_{i_2}^{(4)}), \end{aligned}$$

where the subscripts on the  $B$ 's indicate the sets of  $\xi$ 's still contained in the determinants, and  $\|B^{ij}\| = \|B_{ij}\|^{-1}$ . Similarly,

$$(6.4) \quad \tilde{B}' = \tilde{B}(1 + \tilde{B}_{(1)}^{i_1 i_1} \xi_{i_1}^{(1)} \xi_{i_1}^{(1)})(1 + \tilde{B}_{(1)}^{i_1 i_1} \xi_{i_1}^{(2)} \xi_{i_1}^{(2)})(1 + \tilde{B}_{(12)}^{i_1 i_2} \xi_{i_1}^{(3)} \xi_{i_2}^{(3)})(1 + \tilde{B}_{(12)}^{i_1 i_2} \xi_{i_1}^{(4)} \xi_{i_2}^{(4)}),$$

the inverse now being taken with respect to  $\|\tilde{B}\|$ .

But between, say,  $\tilde{B}_{(12)}^{i_1 i_2}$  and  $B_{(12)}^{i_1 i_2}$ , there is the relation

$$(6.5) \quad \tilde{B}_{(12)}^{i_1 i_2} = B_{(12)}^{i_1 i_2} - B_{(12)}^{i_1 i_2} B_{(12) i_1 i_2} B_{(12)}^{i_1 i_2},$$

where  $\|B_{(12) i_1 i_2}\| = \|B_{(12)}^{i_1 i_2}\|^{-1}$ , that is, the inverse of the matrix obtained by deleting the first four rows and columns of  $\|B_{(12)}^{ij}\|$ . Consequently

$$\tilde{B}_{(12)}^{i_1 i_2} \xi_{i_1}^{(3)} \xi_{i_2}^{(3)} \leq B_{(12)}^{i_1 i_2} \xi_{i_1}^{(3)} \xi_{i_2}^{(3)}$$

with equality holding only for those values of the  $\xi$ 's for which

$$\xi_{i_2}^{(3)} B_{(12)}^{i_1 i_2} = 0 \quad i_2 = 5, \dots, m.$$

And this set of  $\xi$ 's will not make up the entire  $\xi$  space unless  $\|B_{ij}\| = \|\tilde{B}\| + \|B_{i_1 i_2}\|$ . Applying the same kind of reasoning to the other quadratic forms in (6.4), we can therefore show that

$$\begin{aligned} &\int B^{i(N-1)} B'^{-i(N-1)} \tilde{B}'^{-1} d\xi \\ &\leq \tilde{B}^{-1} \int (1 + \tilde{B}_{(1)}^{i_1 i_1} \xi_{i_1}^{(1)} \xi_{i_1}^{(1)})^{-i(N+1)} \dots (1 + \tilde{B}_{(12)}^{i_1 i_2} \xi_{i_1}^{(4)} \xi_{i_2}^{(4)})^{-i(N+1)} d\xi. \end{aligned}$$

The last form can be reduced to a sum of squares with unit coefficients by a linear transformation of the  $\xi^{(4)}$ 's; thus

$$\begin{aligned} &\int B^{i(N-1)} B'^{-i(N-1)} \tilde{B}'^{-1} d\xi \\ (6.6) \quad &\leq \tilde{B}^{-1} \int |\tilde{B}_{(12)}^{i_1 i_2}|^{-1} (1 + \tilde{B}_{(1)}^{i_1 i_1} \xi_{i_1}^{(1)} \xi_{i_1}^{(1)})^{-i(N-1)} \dots (1 + \Sigma \xi_{i_2}^{(4)} \xi_{i_2}^{(4)})^{-i(N+1)} d\xi. \end{aligned}$$

And by making use of the fact that

$$\tilde{B}_{(12)}^{i_1 i_2} = \tilde{B}_{(12)}^{-1} \cdot |B_{(12) i_1 i_2}|,$$

we can express the right-hand side of (6.6) as

$$\check{B}^{-1} \int \check{B}_{(12)}^{i_1 j_1} \cdot |B_{(12) i_1 j_1}|^{-1} (1 + \check{B}^{i_1 j_1} \xi_{i_1}^{(1)} \xi_{j_1}^{(1)})^{-1(N+1)} \dots (1 + \Sigma \xi_{i_2}^{(4)} \xi_{j_2}^{(4)})^{-1(N+1)} d\xi.$$

This in turn becomes [c. f. (6.4)]

$$\begin{aligned} \check{B}^{-1} \int & |B_{(12) i_1 j_1}|^{-1} (1 + \check{B}^{i_1 j_1} \xi_{i_1}^{(1)} \xi_{j_1}^{(1)})^{-1N} (1 + \check{B}_{(1)}^{i_1 j_1} \xi_{i_1}^{(2)} \xi_{j_1}^{(2)})^{-1N} \\ & \times (1 + \check{B}_{(12)}^{i_2 j_2} \xi_{i_2}^{(3)} \xi_{j_2}^{(3)})^{-1N} (1 + \Sigma \xi_{i_2}^{(4)} \xi_{j_2}^{(4)})^{-1(N+1)} d\xi \\ & = \int |B_{(12) i_1 j_1}|^{-1} (1 + \check{B}^{i_1 j_1} \xi_{i_1}^{(1)} \xi_{j_1}^{(1)})^{-1N} (1 + \check{B}_{(1)}^{i_1 j_1} \xi_{i_1}^{(2)} \xi_{j_1}^{(2)})^{-1(N-1)} \\ & \times (1 + \Sigma \xi_{i_2}^{(3)} \xi_{j_2}^{(3)})^{-1N} (1 + \Sigma \xi_{i_2}^{(4)} \xi_{j_2}^{(4)})^{-1(N+1)} d\xi. \end{aligned}$$

At this stage we can write

$$|B_{(12) i_1 j_1}| = |B_{i_1 j_1}| (1 + B^{*i_1 j_1} \xi_{i_1}^{(1)} \xi_{j_1}^{(1)}) (1 + B_{(1)}^{*i_1 j_1} \xi_{i_1}^{(2)} \xi_{j_1}^{(2)}),$$

where  $\|B_{(1)}^{*i_1 j_1}\| = \|B_{(1) i_1 j_1}\|^{-1}$ , and apply the relation

$$B_{(1)}^{*i_1 j_1} = \check{B}_{(1)}^{i_1 j_1} - \check{B}_{(1)}^{i_1 i_2} \check{B}_{(1) i_2 j_2} \check{B}_{(1)}^{j_2 j_1}, \quad \|\check{B}_{(1) i_2 j_2}\| = \|\check{B}_{(1)}^{i_2 j_2}\|^{-1}.$$

Therefore,

$$B_{(1)}^{*i_1 j_1} \xi_{i_1}^{(2)} \xi_{j_1}^{(2)} < \check{B}_{(1)}^{i_1 j_1} \xi_{i_1}^{(2)} \xi_{j_1}^{(2)},$$

unless  $\xi_{i_1}^{(2)} \check{B}_{(1)}^{i_1 i_2} = 0$  ( $i_2 = 3, 4$ ). We can thus continue as follows

$$\begin{aligned} & \int B^{1(N-1)} B^{-1(N-1)} \check{B}^{-1} d\xi \\ & \leq |B_{i_1 j_1}|^{-1} \int (1 + B^{*i_1 j_1} \xi_{i_1}^{(1)} \xi_{j_1}^{(1)})^{-1(N+1)} (1 + B_{(1)}^{*i_1 j_1} \xi_{i_1}^{(2)} \xi_{j_1}^{(2)})^{-1(N+1)} \\ & \quad \times (1 + \Sigma \xi_{i_2}^{(3)} \xi_{j_2}^{(3)})^{-1N} (1 + \Sigma \xi_{i_2}^{(4)} \xi_{j_2}^{(4)})^{-1(N+1)} d\xi. \end{aligned}$$

Transforming the  $\xi^{(2)}$ 's, we get

$$\begin{aligned} & |B_{i_1 j_1}|^{-1} \int |B_{(1)}^{*i_1 j_1}|^{-1} (1 + B^{*i_1 j_1} \xi_{i_1}^{(1)} \xi_{j_1}^{(1)})^{-1(N+1)} (1 + \Sigma \xi_{i_1}^{(2)} \xi_{j_1}^{(2)})^{-1(N+1)} \\ & \quad \times (1 + \Sigma \xi_{i_2}^{(3)} \xi_{j_2}^{(3)})^{-1N} (1 + \Sigma \xi_{i_2}^{(4)} \xi_{j_2}^{(4)})^{-1(N+1)} d\xi. \end{aligned}$$

Since  $|B_{(1)}^{*i_1 j_1}|^{-1} = |B_{(1) i_1 j_1}|$ , this becomes

$$\begin{aligned} & |B_{i_1 j_1}|^{-1} \int (1 + B^{*i_1 j_1} \xi_{i_1}^{(1)} \xi_{j_1}^{(1)})^{-1N} (1 + \Sigma \xi_{i_1}^{(2)} \xi_{j_1}^{(2)})^{-1(N+1)} \\ & \quad \times (1 + \Sigma \xi_{i_2}^{(3)} \xi_{j_2}^{(3)})^{-1N} (1 + \Sigma \xi_{i_2}^{(4)} \xi_{j_2}^{(4)})^{-1(N+1)} d\xi \\ & = \int (1 + \Sigma \xi_{i_1}^{(1)} \xi_{j_1}^{(1)})^{-1N} (1 + \Sigma \xi_{i_1}^{(2)} \xi_{j_1}^{(2)})^{-1(N+1)} \\ & \quad \times (1 + \Sigma \xi_{i_2}^{(3)} \xi_{j_2}^{(3)})^{-1N} (1 + \Sigma \xi_{i_2}^{(4)} \xi_{j_2}^{(4)})^{-1(N+1)} d\xi. \end{aligned}$$

Collecting these results, we finally obtain

$$\begin{aligned}
 (6.7) \quad & K_1 \int B^{i(N-1)} B^{r-1(N-1)} \tilde{B}^{-1} d\xi \\
 & \leq K_1 \int (1 + \sum \xi_{i_1}^{(1)} \xi_{i_1}^{(1)})^{-1N} (1 + \sum \xi_{i_1}^{(2)} \xi_{i_1}^{(2)})^{-1(N+1)} \\
 & \quad \times (1 + \sum \xi_{i_1}^{(3)} \xi_{i_1}^{(3)})^{-1N} (1 + \sum \xi_{i_1}^{(4)} \xi_{i_1}^{(4)})^{-1(N+1)} d\xi
 \end{aligned}$$

with equality only in case  $H_I$  is true. But the right side of (6.7) is the first moment of  $\lambda_I^{2/N}$  computed under the hypothesis  $H_I$ , while the left side gives the corresponding moment in the general case.

The possibility of carrying out this reduction for the case in which the matrix  $\|\tilde{B}\|$  has more than two blocks, or blocks of unequal size, seems sufficiently clear. And to obtain higher moments, we have only to introduce the proper number of  $\xi$ 's into each set. We then have:

**THEOREM IIIa.** *Let  $\lambda_i$  be the likelihood ratio appropriate to testing the hypothesis  $H_I$  that the normally distributed variates  $x^1, \dots, x^m$  fall into the mutually independent sets  $x^1, \dots, x^{m_1}; \dots; x^{m_{p-1}+1}, \dots, x^m$ . Then the expected value of  $(\lambda_I^{2/N})^h$ ,  $h = \frac{1}{2}, 1, 1\frac{1}{2}, \dots$ , is greater under the null hypothesis  $H_I$  than under any alternative hypothesis in  $\Omega$ .*

**7. The general regression problem.** Let the variates  $x^1, \dots, x^t$  be distributed according to the law

$$(7.1) \quad \frac{|B_{ij}|^{\frac{1}{2}}}{\pi^{\frac{1}{2}t}} e^{-B_{ij}(x^i - C_p^i x^u - C_p^i x^v)(x^i - C_p^i x^v - C_p^i x^v)}$$

Throughout this section, let the ranges of the indices be

$$\begin{aligned}
 i, j &= 1, \dots, t & p, q &= t+1, \dots, m \\
 r, s &= 1, \dots, m & r', s' &= 1', \dots, t+q \\
 \mu, \nu &= t+1, \dots, t+q & \sigma, \tau &= t+q+1, \dots, m.
 \end{aligned}$$

In (7.1) we therefore have  $t$  random variates, and  $m - t$  fixed variates. Consider the hypothesis  $H$  that the  $x^i$  are independent of the last set of  $x$ 's, namely  $x^r$ . We have

$$\Omega: \|B_{ij}\| \text{ positive definite, } -\infty < C_p^i < \infty,$$

while for  $\omega$  we impose the additional requirement

$$C_p^i = 0.$$

Thus in general we have for the distribution of random samples  $0_N$ ,  $N \geq m$ ,

$$(7.2) \quad \frac{|B_{ij}|^{\frac{1}{2}N}}{\pi^{\frac{1}{2}N}} e^{-\sum_{i=1}^N B_{ij}(x_i^i - C_p^i x_i^u - C_p^i x_i^v)(x_i^i - C_p^i x_i^v - C_p^i x_i^v)},$$

while when  $H$  is true, we have

$$(7.3) \quad P = \frac{|B_{ij}|^{1N}}{\pi^{1Nt}} e^{-\sum_{\alpha=1}^N B_{ij}(x_\alpha^i - C_p^i x_\alpha^p)(x_\alpha^j - C_q^j x_\alpha^q)}.$$

Differentiating (7.2) with respect to the  $B$ 's and  $C$ 's and setting the derivatives equal to zero gives us the conditions

$$(7.4) \quad \sum_{\alpha=1}^N C_p^i x_\alpha^p x_\alpha^q = \sum_{\alpha=1}^N x_\alpha^i x_\alpha^q,$$

$$(7.5) \quad B^{ij} = \frac{2}{N} \sum_{\alpha=1}^N (x_\alpha^i - C_p^i x_\alpha^p)(x_\alpha^j - C_q^j x_\alpha^q).$$

As in section 2, we put

$$a^{rs} = \sum_{\alpha=1}^N x_\alpha^r x_\alpha^s.$$

and assume that the fixed values  $x_\alpha^p$  have been so chosen that  $\|a^{pq}\|$  is positive definite. Then (7.4) and (7.5) can be combined to give

$$B^{ij} = \frac{2}{N} (a^{ij} - a^{ip} a_{pq}' a^{qj}) = \frac{2}{N} \tilde{a}^i$$

where  $\|a_{pq}'\|^{-1} = \|a^{pq}\|$ . It then follows that

$$P_\Omega = |\tilde{a}^{ij}|^{-1N} \left(\frac{N}{2\pi}\right)^{1Nt} e^{-1N}.$$

Similarly,

$$P_\omega = |\tilde{a}_0^{ij}|^{-1N} \left(\frac{N}{2\pi}\right)^{1Nt} e^{-1N},$$

where

$$\tilde{a}_0^{ij} = a^{ij} - a^{i\mu} a_{\mu\nu}' a^{\nu j}, \quad \|a_{\mu\nu}'\|^{-1} = \|a^{\mu\nu}\|.$$

The matrix  $\|a^{rs}\|$  will be positive definite except for a set of probability zero, so that we can consider  $\|\tilde{a}^{ij}\|$  as the inverse of the matrix obtained by removing the last  $m - t$  rows and columns of the inverse of  $\|a^{rs}\|$ , and  $\|\tilde{a}_0^{ij}\|$  as the inverse of the matrix obtained by removing the last  $q$  rows and columns of  $\|a^{r's'}\|^{-1}$ . Then by the Jacobi theorem

$$|\tilde{a}^{ij}|^{-1} = \frac{|a^{pq}|}{|a^{rs}|}, \quad |\tilde{a}_0^{ij}|^{-1} = \frac{|a^{\mu\nu}|}{|a^{r's'}|}$$

so that the appropriate likelihood ratio is given by

$$\lambda^{2/N} = \frac{|a^{rs}|}{|a^{r's'}|} \cdot \frac{|a^{\mu\nu}|}{|a^{pq}|}.$$

It will be advantageous to complete the matrix  $\|B_{ij}\|$  in (7.1) by defining

$$(7.6) \quad \begin{aligned} B_{ip} &= -B_{ij}C_p^j, \\ B_{pq} &= C_p^i B_{ij}C_q^j. \end{aligned}$$

(Evidently  $B_{ip} = 0$  for  $i = 1, \dots, t$  and fixed  $p$ , if and only if  $C_p^j = 0$ ,  $j = 1, \dots, t$ ). We can now write (7.2) as

$$(7.7) \quad P(x, B) = \frac{|B_{ij}|^{1N}}{\pi^{1Nt}} e^{-\sum_{a=1}^N B_{ij}(x_a^i + B^{ik} B_{kp} x_a^p)(x_a^j + B^{jl} B_{lq} x_a^q)}.$$

We next notice that  $\lambda$  is invariant under the transformations

$$x^i \rightarrow \alpha_j^i x^j, \quad x^r \rightarrow \beta_r^s x^s,$$

so that if we put

$$I(B, \lambda_s) = \int_S P(x, B) dx_1^1 \dots dx_N^t,$$

where the integral is extended over the region

$$S: 0 \leq \lambda \leq \lambda_s,$$

it turns out that

$$I(B, \lambda_s) \equiv I(B^*, \lambda_s),$$

provided

$$B_{ij}^* = \alpha_i^k B_{kl} \alpha_j^l, \quad B_{i\mu}^* = \alpha_i^k B_{k\mu}, \quad B_{i\sigma}^* = \alpha_i^k B_{k\tau} \beta_\sigma^\tau.$$

To prove the locally unbiased character of the test, we may therefore consider the derivatives

$$\frac{\partial}{\partial B_{i\sigma}^*} I(B_0^*, \lambda_s), \quad \frac{\partial^2}{\partial B_{i\sigma}^* \partial B_{j\tau}^*} I(B_0^*, \lambda_s),$$

and assume that  $\|B_{ij}^*\|$  and  $\|a^{\mu\nu}\|$  are in diagonal form. We also observe that  $\lambda$  is unaltered by the transformation

$$x^i \rightarrow x^i + B^{ik} B_{k\mu} x^\mu.$$

We therefore have

$$I(B^*, \lambda_s) = \frac{|B_{ij}^*|^{1N}}{\pi^{1Nt}} \int_S e^{-\sum_{a=1}^N B_{ij}^*(x_a^i + B^{ik} B_{k\sigma}^* x_a^\sigma)(x_a^j + B^{jl} B_{l\tau}^* x_a^\tau)} dx.$$

Thus,

$$\frac{\partial}{\partial B_{i\sigma}^*} I(B_0^*, \lambda_s) = -2 \frac{|B_{ij}^*|^{1N}}{\pi^{1Nt}} \int_S \sum_{a=1}^N x_a^k x_a^\sigma e^{-\sum_{a=1}^N B_{ij}^*(x_a^i + B^{ik} B_{k\sigma}^* x_a^\sigma)(x_a^j + B^{jl} B_{l\tau}^* x_a^\tau)} dx,$$

which is easily seen to be zero. Again, consider a non-repeated second partial derivative, say

$$\frac{\partial^2}{\partial B_{ks}^* \partial B_{lr}^*} I(B_0^*, \lambda_s) \\ = -2 \frac{|B_{ij}^*|^{1N}}{\pi^{1Nl}} \int_s \left( B^{*kl} a^{\sigma\tau} - 2 \sum_{\alpha=1}^N x_\alpha^\sigma x_\alpha^k \cdot \sum_{\beta=1}^N x_\beta^\tau x_\beta^l \right) e^{-\sum_{a=1}^N x_a^\sigma x_a^i x_a^\tau x_a^j} d\tau.$$

This plainly vanishes if  $k \neq l$ ; but it is by no means easy to see what happens when  $k = l$ , even when  $\sigma \neq \tau$ . Let us therefore study the distribution law of  $\lambda^{2/N}$  for the case,

$$B_{is} = 0, \quad i \neq 1.$$

(We shall not, however, assume that the transformation  $B \rightarrow B^*$  has been made on the  $B$ 's.)

Define

$$\tilde{B}_{pq} = B_{pq} - B_{pi} B^{ij} B_{jq}, \\ \tilde{a}^{\sigma\tau} = a^{\sigma\tau} - a^{\sigma\mu} a_{\mu\nu} a^{\nu\tau},$$

where  $\|a_{\mu\nu}\|$  now stands for the inverse of  $\|a^{\mu\nu}\|$ . These expressions will arise when we adapt Wilks' method of moment generating operators [13], based on the identity

$$(7.8) \quad \int e^{-B_{rs} a^{\sigma\tau}} dx_1^1 \dots dx_N^l = \pi^{1Nl} B^{-1N} \exp(-\tilde{B}_{pq} a^{pq})$$

to the problem. We shall understand from now on that  $B = |B_{ij}|$  and  $\|B^{ij}\| = \|B_{ij}\|^{-1}$ . Let us rearrange the form in the exponential on the right, thus:

$$\tilde{B}_{pq} a^{pq} = (\tilde{B}_{\mu\nu} a^{\mu\nu} + 2B_{\mu\sigma} a^{\mu\sigma} + B_{\sigma\tau} a^{\sigma\tau} - 2B_{\mu i} B^{ij} B_{j\sigma} a^{\mu\sigma} \\ - B_{\sigma i} B^{ij} B_{j\tau} a^{\sigma\mu} a_{\mu\nu} a^{\nu\tau}) - B_{\sigma i} B^{ij} B_{j\tau} \tilde{a}^{\sigma\tau} \\ = Q - B_{\sigma i} B^{ij} B_{j\tau} \tilde{a}^{\sigma\tau} \\ = Q - B^{ij} y_{ij}.$$

A subscript  $\beta$  will denote the result of replacing  $B_{r's'}$  by  $B_{r's'} + \beta_{r's'}$ , and a prime will indicate that each  $\beta_{r's'}$  has been replaced by  $\beta_{r's'} + \xi_{r'} \xi_{s'}$ . Consider now the result of integrating the right hand side of (7.8) after these replacements have been made:

$$(7.9) \quad \pi^{1Nl} \int B_\beta'^{-1N} \exp(-\tilde{B}'_{pq\beta} a^{pq}) d\xi_1 \dots d\xi_l d\xi_{l+1} \dots d\xi_{l+q} \\ \pi^{1Nl} \int B_\beta'^{-1N} e^{B_\beta'^{ij} y_{ij}} \left( \int e^{-Q_\beta} d\xi_\mu \right) d\xi_l,$$

Let us integrate first with respect to the  $\xi_\mu$ . Wilks has shown how to write  $Q'_\beta$  in the form

$$Q'_\beta = -Q'_{1\beta} + B_{\rho\sigma\beta} a^{\rho\sigma} + \frac{B_\beta}{B'_\beta} a^{\mu\nu} \xi_\mu \xi_\nu - 2B_{\rho i\beta} B_\beta'^{ij} a^{\rho\nu} \xi_i \xi_\nu,$$

where

$$Q'_{1\beta} = B_{\mu i\beta} B_\beta'^{ij} B_{j\nu\beta} a^{\mu\nu} + 2B_{\mu i\beta} B_\beta'^{ij} B_{j\sigma} a^{\mu\sigma} + B_{\sigma i} B_\beta'^{ij} B_{j\tau} a^{\sigma\tau} a_{\mu\nu} a^{\mu\nu}.$$

This latter expression is thus free of the  $\xi_\mu$ . Consequently,

$$\int e^{-Q'_\beta} d\xi_\mu = \left( \frac{B'_\beta}{B_\beta} \right)^{1/2} |a^{\mu\nu}|^{-1} \pi^{1/2} e^{-B_{\rho\sigma\beta} a^{\rho\sigma}} e^{Q'_{1\beta} + Q'_{2\beta}},$$

where

$$Q'_{2\beta} = \frac{B'_\beta}{B_\beta} a_{\mu\nu} (a^{\mu\rho} B_{\rho i\beta} B_\beta'^{ij} \xi_i) (a^{\nu\sigma} B_{\sigma k\beta} B_\beta'^{kl} \xi_k),$$

which can be written

$$\begin{aligned} \frac{B'_\beta}{B_\beta} \{ & B_{\mu i\beta} B_\beta'^{ij} B_{\nu k\beta} B_\beta'^{kl} \xi_i \xi_k a^{\mu\nu} + 2B_{\mu i\beta} B_\beta'^{ij} B_{\sigma k\beta} B_\beta'^{kl} \xi_i \xi_k a^{\mu\sigma} \\ & + B_{\sigma i} B_\beta'^{ij} B_{\tau k\beta} B_\beta'^{kl} \xi_i \xi_k a^{\sigma\tau} a_{\mu\nu} a^{\mu\nu} \}. \end{aligned}$$

The method of reduction used by Wilks can now be applied to  $Q'_{1\beta}$  and  $Q'_{2\beta}$ , and gives

$$Q'_{1\beta} + Q'_{2\beta} = B_{\mu i\beta} B_\beta'^{ij} B_{j\nu\beta} a^{\mu\nu} + 2B_{\mu i\beta} B_\beta'^{ij} B_{j\sigma} a^{\mu\sigma} + B_{\sigma i} B_\beta'^{ij} B_{j\tau} a^{\sigma\tau} a_{\mu\nu} a^{\mu\nu},$$

an expression which does not involve the  $\xi$ 's. Thus

$$(7.10) \quad \int e^{-Q'_\beta} d\xi_\mu = \pi^{1/2} |a^{\mu\nu}|^{-1} B_\beta^{-1/2} e^{-Q_\beta} B_\beta'^{1/2}.$$

Now the quantity

$$e^{B_\beta'^{ij} y_{ij}} = \sum_{\nu=0}^{\infty} \frac{(y_{ij} B_\beta'^{ij})^\nu}{\nu!} B_\beta'^{-\nu},$$

where  $B^{ij}$  stands for the cofactor of  $B_{ij}$  in  $||B_{ij}||$ , can be expressed in terms of  $B'_\beta$ , provided we use our assumption that  $B_{i\sigma} = 0$ ,  $i \neq 1$ , whereupon  $y_{ij} B_\beta'^{ij}$  reduces to the single term  $y B_\beta'^{11}$ . In fact, we have

$$\begin{aligned} (7.11) \quad E[g_\beta | a^{\rho\sigma}] &= \bar{K} \prod_{i=1}^h \psi(N - m + t + 1 - i, 2h) |a^{\rho\sigma}|^h \pi^{1/2} B_\beta^{-1/2(N+h)} \\ &\times \exp(-\bar{B}_{\rho\sigma\beta} a^{\rho\sigma}) = \bar{K} \prod_{i=1}^h \psi \cdot \pi^{1/2} |a^{\rho\sigma}|^h B_\beta^{-1/2} e^{-Q_\beta} \sum_{\nu=1}^{\infty} \frac{(y B_\beta'^{11})^\nu}{\nu!} B_\beta^{-1/2(N-\nu)+h+\nu}, \end{aligned}$$



where, following the notation used by Wilks [13],

$$g_\beta = e^{-\beta_{rs} a^{rs}}, \quad \bar{K} = \pi^{-iNt} B^{iN} \exp(-\bar{B}_{pq} a^{pq}),$$

$$\psi(a, b) = \frac{\Gamma[\frac{1}{2}(a+b)]}{\Gamma[\frac{1}{2}a]}.$$

And (7.11) can be written as

$$(7.12) \quad E[g_\beta | a^{rs} |^h] = \bar{K} \pi^{iNt} \prod_{i=1}^t \psi \cdot |a^{pq}|^h B_\beta^{-iq} e^{-q_\beta} \\ \times \sum_{\nu=0}^{\infty} \frac{y^\nu}{\nu!} \frac{\Gamma[\frac{1}{2}(N-q)+h]}{\Gamma[\frac{1}{2}(N-q)+h+\nu]} \frac{\partial^\nu}{\partial u^\nu} (B_{\beta u}^{-\{\frac{1}{2}(N-q)+h\}})_{u=0},$$

where  $B_u$  stands for the result of replacing  $B_{11}$  by  $B_{11} - u$ . Changing  $\beta_{r's'}$  into  $\beta_{r's'} + \xi_{r'} \xi_{s'}$  and integrating, we then find that by virtue of (7.10)

$$(7.13) \quad E[g_\beta | a^{rs} |^h | a^{r's'} |^{-i}] = \bar{K} \pi^{iNt} \prod_{i=1}^t \psi \cdot |a^{pq}|^h |a^{\mu\nu}|^{-i} B_\beta^{-iq} e^{-q_\beta} \\ \times \pi^{-i} \prod_{\nu=0}^{\infty} \frac{y^\nu}{\nu!} \frac{\Gamma[\frac{1}{2}(N-q)+h]}{\Gamma[\frac{1}{2}(N-q)+h+\nu]} \frac{\partial^\nu}{\partial u^\nu} \int B_{\beta u}^{-\{\frac{1}{2}(N-q)+h\}} d\xi_i \Big]_{u=0}.$$

Now

$$\int B_{\beta u}^{-\{\frac{1}{2}(N-q)+h\}} d\xi_i = B_{\beta u}^{-\{\frac{1}{2}(N-1-q)+h\}} \pi^{it} \prod_{i=1}^t \psi(N-q+2h+1-i, -1),$$

so that (7.13) becomes

$$(7.14) \quad E[g_\beta | a^{rs} |^h | a^{r's'} |^{-i}] = \bar{K} \pi^{iNt} \prod_{i=1}^t \psi(N-m+t+1-i, 2h) \\ \times \prod_{i=1}^t \psi(N-q+2h+1-i, -1) |a^{pq}|^h |a^{\mu\nu}|^{-i} \\ \times B_\beta^{-iq} e^{-q_\beta} \sum_{\nu=0}^{\infty} \frac{y^\nu}{\nu!} \frac{\Gamma[\frac{1}{2}(N-q)+h]}{\Gamma[\frac{1}{2}(N-q)+h+\nu]} \frac{\partial^\nu}{\partial u^\nu} (B_{\beta u}^{-\{\frac{1}{2}(N-1-q)+h\}})_{u=0}.$$

Comparing (7.14) with (7.12), and making use of the fact that

$$\psi(a, -1)\psi(1-1, -1) \dots \psi(a-2h+1, -1) = \psi(a, -2h),$$

we thus have

$$E[g_\beta | a^{rs} |^h | a^{r's'} |^{-h}] = \bar{K} \pi^{iNt} \prod_{i=1}^t \psi(N-m+t+1-i, 2h) \\ \times \prod_{i=1}^t \psi(N-q+2h+1-i, -2h) |a^{pq}|^h |a^{\mu\nu}|^{-h} B_\beta^{-iq} e^{-q_\beta} \\ \times \sum_{\nu=0}^{\infty} \frac{y^\nu}{\nu!} \frac{\Gamma[\frac{1}{2}(N-q)+h]}{\Gamma[\frac{1}{2}(N-q)+h+\nu]} \frac{\partial^\nu}{\partial u^\nu} [B_{\beta u}^{-\frac{1}{2}(N-q)}]_{u=0}.$$

Setting the  $\beta$ 's equal to zero, performing the differentiation, and recalling the definitions of  $\tilde{K}$  and  $Q_\beta$ , we then find

$$(7.15) \quad E[(\lambda^{2/N})^h] = \prod_{i=1}^t \psi(N - m + t + 1 - i, 2h) \prod_{i=1}^t \psi(N - q + 2h + 1 - i, -2h) \\ \times e^{-\nu B^{11}} \sum_{\nu=0}^{\infty} \frac{(yB^{11})^\nu}{\nu!} \frac{\Gamma[\frac{1}{2}(N - q) + h]}{\Gamma[\frac{1}{2}(N - q) + h + \nu]} \frac{\Gamma[\frac{1}{2}(N - q) + \nu]}{\Gamma[\frac{1}{2}(N - q)]}.$$

Taking the first factor from each product, we can convert (7.15) into

$$\prod_{i=2}^t \psi(N - m + t + 1 - i, 2h) \prod_{i=2}^t \psi(N - q + 2h + 1 - i, -2h) \\ \times e^{-\nu B^{11}} \sum_{\nu=0}^{\infty} \frac{(yB^{11})^\nu}{\nu!} \frac{\Gamma[\frac{1}{2}(N - m + t) + h]}{\Gamma[\frac{1}{2}(N - m + t)]} \frac{\Gamma[\frac{1}{2}(N - q) + \nu]}{\Gamma[\frac{1}{2}(N - q) + h + \nu]}.$$

This last product of ratios of  $\Gamma$ 's is equivalent to

$$\frac{\Gamma[\frac{1}{2}(N - q) + \nu]}{\Gamma[\frac{1}{2}(N - m + t)]\Gamma[\frac{1}{2}(m - t - q) + \nu]} \cdot \frac{\Gamma[\frac{1}{2}(m - t - q) + \nu]\Gamma[\frac{1}{2}(N - m + t) + h]}{\Gamma[\frac{1}{2}(N - q) + h + \nu]}.$$

Thus the moments of  $\lambda^{2/N}$  are connected with an integral equation of type B [12] and  $\lambda^{2/N}$  is distributed like the product

$$z \cdot \theta_2 \dots \theta_t, \quad 0 \leq z \leq 1, 0 \leq \theta_i \leq 1,$$

where the joint distribution of the  $\theta$ 's is

$$f(\theta) = \prod_{i=2}^t \frac{\Gamma[\frac{1}{2}(N - q + 1 - i)]}{\Gamma[\frac{1}{2}(N - m + t + 1 - i)]\Gamma[\frac{1}{2}(m - t - q)]} \cdot \theta_1^{(N - m + t + 1 - i) - 1} (1 - \theta_i)^{i(m - t - q) - 1},$$

and  $z$  is distributed independently of the  $\theta$ 's with the distribution

$$(7.16) \quad F(z) = e^{-\nu B^{11}} \sum_{\nu=0}^{\infty} \frac{(yB^{11})^\nu}{\nu!} \frac{z^{i(N - m + t) - 1} (1 - z)^{i(m - t - q) + \nu - 1}}{B[\frac{1}{2}(N - m + t), \frac{1}{2}(m - t - q) + \nu]}.$$

The probability that  $0 \leq \lambda \leq \lambda_*$  is therefore

$$I(y, \lambda_*) = \int_S f(\theta) F(z) dz d\theta_2 \dots d\theta_t,$$

where  $S$  is the region  $0 \leq \theta_2 \dots \theta_t, z < \lambda_*^{2/N}$ . Putting  $\varphi(\theta)$  for the upper limit of  $z$  in  $S$  for fixed  $\theta$ , and  $S_\theta$  for the projection of  $S$  into the  $\theta$  space, we then have

$$I(y, \lambda_0) = \int_{S_\theta} f(\theta) \left\{ e^{-\nu B^{11}} \sum_{\nu=0}^{\infty} \frac{(yB^{11})^\nu}{\nu!} \int_0^\varphi \frac{z^{i(N - m + t) - 1} (1 - z)^{i(m - t - q) - 1 + \nu}}{B[\frac{1}{2}(N - m + t), \frac{1}{2}(m - t - q) + \nu]} dz \right\} d\theta.$$

If we replace  $z$  by  $(1 - z)$  we then find

$$(7.17) \quad I(y, \lambda_0) = \int_{S_\theta} f(\theta) \\ \times \left\{ e^{-\nu B^{11}} \sum_{\nu=0}^{\infty} \frac{(yB^{11})^\nu}{\nu!} \frac{B[\frac{1}{2}(m - t - q) + \nu, \frac{1}{2}(N - m + t); 1 - \varphi]}{B[\frac{1}{2}(m - t - q) + \nu, \frac{1}{2}(N - m + t)]} \right\} d\theta.$$

As far as  $y$  is concerned, (7.17) is essentially the same as (2.8). The computation which was made there, together with the type of reasoning employed in the latter part of section 5 in connection with the independence test for several blocks, then shows that

$$\frac{\partial}{\partial y} I(y, \lambda_i) > 0 \quad (0 < \epsilon < 1).$$

Remembering that

$$y = \tilde{a}^{\sigma\tau} B_{\sigma 1} B_{\tau 1},$$

we see that

$$\left( \frac{\partial y}{\partial B_{\sigma 1}} \right)_0 = 0, \quad \frac{\partial^2 y}{\partial B_{\sigma 1} \partial B_{\tau 1}} = 2\tilde{a}^{\sigma\tau},$$

and we remark that the assumed positive definiteness of  $\|\tilde{a}^{pq}\|$  implies that of  $\|\tilde{a}^{\sigma\tau}\|$ . Hence the relation

$$\left( \frac{\partial^2}{\partial B_{1\sigma} \partial B_{1\tau}} I(y, \lambda_i) \right)_0 = \left( \frac{\partial I}{\partial y} \right)_0 \tilde{a}^{\sigma\tau}$$

together with the fact that we could have obtained the analogue of (7.17) under the assumption

$$B_{i\sigma} = 0 \quad i \neq i_0,$$

where  $i_0$  is any fixed number in the set  $1, \dots, t$ , shows that the matrix of second partial derivatives is positive definite when  $H$  is true.

Thus we have

**THEOREM IV.** *Let  $x^1, \dots, x^t$  be normally distributed about means which are linear functions of certain fixed variates  $x^{t+1}, \dots, x^m$ . Then the likelihood ratio test for the hypothesis that the distribution of  $x^1, \dots, x^t$  depends only on a selected subset  $x^{t+1}, \dots, x^{t+q}$  of the fixed variates is locally unbiased.*

The result of this section has its most immediate application to those problems in the analysis of variance which require simultaneous consideration of several interrelated dependent variables  $x^1, \dots, x^t$  in conjunction with a given set of independent variables  $x^{t+1}, \dots, x^m$  [15]. For the usual hypothesis to be tested in this case is that  $x^1, \dots, x^t$  are jointly independent of, say,  $x^{t+q+1}, \dots, x^m$ .

To return to the general case of (7.1), the method of this section can also be used to test the hypothesis that the regression coefficients referring to the  $x^\sigma$  have particular values, say

$$C_\sigma^i = C_{\sigma 0}^i \quad i = 1, \dots, t; \sigma = t + q + 1, \dots, m,$$

the remaining  $C$ 's and the  $B$ 's being left unspecified. Since we have

$$x^i - C_\mu^i x^\mu - C_\sigma^i x^\sigma = x^i - C_\mu^i x^\mu - (C_\sigma^i - C_{\sigma 0}^i) x^\sigma - C_{\sigma 0}^i x^\sigma,$$

by the device of replacing  $x'_a$  by  $x'_a - C'_{a0}x'_a$ , we can reduce this problem to that of testing the hypothesis that

$$C'^i = C'_i - C'_{i0} = 0.$$

Similarly, the problem of testing whether the linear functions  $u'_i = \alpha'_i C'_i$  have specified values  $u'_{i0}$  comes under the same heading [7].

A particularly interesting case of the general regression problem is that in which  $m = t + q + 1$ , so that the null hypothesis  $H$  states that the chance variables  $x^i$  are independent of the fixed variate  $x^m$ , though they may depend upon  $x^{t+1}, \dots, x^{m-1}$ . In this case we are able to find the exact distribution law of  $\lambda^{2/N}$  without assuming that any of the regression coefficients  $C^i$  are zero. For the quantity

$$(7.18) \quad \sum_{\nu=0}^{\infty} \frac{(y_{ij} B^{ij})^{\nu}}{\nu!} B_{\beta}^{-[\frac{1}{2}(N-q)+h+\nu]},$$

which would have occurred in (7.11) had it not been for the restriction  $B_{i\alpha} = 0$  ( $i \neq 1$ ), can now be expressed in terms of  $B_{\beta}$  even without this restriction. By definition

$$y_{ij} B^{ij} = \tilde{a}^{mm} B^{ij} B_{mi} B_{mj}$$

and the vanishing of the  $B_{mi}$  is equivalent to the vanishing of the regression coefficients  $C_m^i$  associated with  $x^m$ . And since

$$|B_{ij} - u \tilde{a}^{mm} B_{mi} B_{mj}| = B - u \tilde{a}^{mm} B^{ij} B_{mi} B_{mj},$$

we can write (7.18) in the form

$$\sum_{\nu=0}^{\infty} \frac{1}{\nu!} \frac{\Gamma[\frac{1}{2}(N-q)+h]}{\Gamma[\frac{1}{2}(N-q)+h+\nu]} \cdot \frac{\partial^{\nu}}{\partial u^{\nu}} [B_{\beta u}^{-[\frac{1}{2}(N-q)+h]}]_{u=0},$$

where

$$||B_{\beta u}|| = ||B_{ij\beta} - u \tilde{a}^{mm} B_{mi} B_{mj}||$$

is positive definite provided  $u$  is sufficiently small. Thus the moments of  $\lambda^{2/N}$  can be found from (7.15) if we put  $\tilde{a}^{mm} B^{ij} B_{mi} B_{mj} = y_{ij} B^{ij}$  in place of  $y B^{11}$ . Moreover, it can be seen that when the value  $m = t + q + 1$  is substituted into (7.15), that expression reduces to

$$E[(\lambda^{2/N})^h] = e^{-y_{ij} B^{ij}} \sum_{\nu=0}^{\infty} \frac{(y_{ij} B^{ij})^{\nu}}{\nu!} \frac{B[\frac{1}{2}(N-m+1)+h, \frac{1}{2}(m-q-1)+\nu]}{B[\frac{1}{2}(N-m+1), \frac{1}{2}(m-q-1)+\nu]}$$

so that  $\lambda^{2/N}$  is distributed like  $w$ , where

$$(7.19) \quad f(w) = e^{-y_{ij} B^{ij}} \sum_{\nu=0}^{\infty} \frac{(y_{ij} B^{ij})^{\nu}}{\nu!} \frac{w^{\frac{1}{2}(N-m+1)-1} (1-w)^{\frac{1}{2}(m-q-1)-1+\nu}}{B[\frac{1}{2}(N-m+1), \frac{1}{2}(m-q-1)+\nu]}.$$

The distribution law of  $\lambda^{2/N}$  for this case is thus closely related to that obtained in the treatment of the regression problem with one dependent variate in section 2. Applying the argument used there, we can obtain:

**THEOREM IVa.** *The likelihood ratio test for the hypothesis that in a population of the type (7.1) the variates  $x^i$  are independent of  $x^m$ —the case  $m = t + q + 1$  of Theorem IV—is completely unbiased.*

If we specialize the problem somewhat further, considering the case  $q = 0$ ,  $x_a^m = 1$  (so that  $m = t + 1$ ), we find that the likelihood ratio takes the form

$$\lambda^{2/N} = \frac{1}{1 + N v_{ij} \bar{x}^i \bar{x}^j} = \frac{1}{1 + T},$$

where  $v^{ij} = \sum_{\alpha=1}^N (x_a^i - \bar{x}^i)(x_a^j - \bar{x}^j)$ , and  $T$  is Hotelling's generalization [5] of Student's ratio. In this case we are testing the hypothesis that the  $x^i$  are distributed with zero means. The exact distribution law of

$$T = \frac{1 - \lambda^{2/N}}{\lambda^{2/N}}$$

was recently published by P. L. Hsu [6], who obtained it in a very elegant fashion by means of the Laplace transform. He has also shown that the resulting test is *most powerful* in the sense that, of all critical regions  $S$  for which

$$P\{x \subset S\} = \epsilon + \frac{1}{2}\alpha B^{ij} b_i b_j + R(b)$$

(where  $\epsilon$  and  $\alpha$  are independent of the  $B^{ij}$  and of the means  $b_i$ , and  $R$  is an infinitesimal of at least the third order as all  $b_i$  tend to zero), the critical region defined by

$$S: T \geq T_0$$

has the largest possible value of  $\alpha$ . Tang's tables [11] make it evident that this largest possible value of  $\alpha$  is actually positive and that the test is in fact unbiased for all values of the  $b$ 's when  $\epsilon = .05$  or  $\epsilon = .01$ . The results of this section may be used to show that this property extends to all probability levels other than  $\epsilon = 0$  and  $\epsilon = 1$ .

The application of Hotelling's  $T$  is by no means confined to the above case. Other hypotheses which can be tested by means of this statistic are discussed by Hsu [6]. In addition it is now known that the Studentized  $D^2$ , devised by Mahalanobis for measuring the "distance" between two normal multivariate populations, is proportional to Hotelling's  $T$ . This fact is pointed out by R. C. Bose and N. Roy [1], who have obtained the exact distribution of  $D^2$  for the case in which the two populations from which the samples are drawn are assumed to have the same matrix of variances and covariances, but are allowed to have different sets of means; their work, however, is quite independent of Hsu's. They also note that  $D^2$  is proportional to the ratio which arises in Fisher's method of multiple measurements [4].

**8. Summary.** The method of likelihood-ratios is of practical as well as theoretical importance, because it provides a unified approach to the problem of testing statistical hypotheses. In this paper we have investigated many of the tests which this method yields when applied to hypotheses about sets of regression coefficients and covariances in normal populations. By studying the probability functions of the corresponding  $\lambda$ -criteria we are able to show that these tests are "good," in the sense that they are unbiased even for small samples.

Among the completely unbiased tests which can be based on the likelihood-ratio method, our discussion includes: the multiple correlation coefficient, with or without fixed variates [13]; Hotelling's generalized  $T$  test [6] and the statistically equivalent "Studentized  $D^2$ " [1]; the ordinary analysis of variance and covariance for orthogonal or non-orthogonal data [11, 16], as well as related tests of linear hypotheses in the case of one chance variable.

With respect to the analysis of variance for two or more variables [15] and certain other hypotheses regarding regression coefficients in multivariate populations, though there are indications that the tests are completely unbiased, we have succeeded in demonstrating this property only in the local sense.

Finally, the likelihood-ratio test for the hypothesis that the variates fall into certain specified mutually independent sets [14] is shown to be unbiased, at least locally, and has the additional property described in Theorem IIIa.

In conclusion, much more than a word of acknowledgment is due to Professor S. S. Wilks of Princeton University, to whom the writer is greatly indebted for advice and encouragement.

## REFERENCES

- [1] R. C. BOSE AND N. ROY, "The Distribution of the Studentized  $D^2$ -statistic," *Sankhyā*, Vol. 4 (1938-39), pp. 19-38.
- [2] G. W. BROWN, "On the Power of the  $L_1$  Test for Equality of Several Variances," *Annals of Math. Stat.*, Vol. 10 (1939), p. 124.
- [3] R. A. FISHER, "The General Sampling Distribution of the Multiple Correlation Coefficient," *Proc. Roy. Soc. Lond.*, series A, Vol. 121 (1928), pp. 196-203.
- [4] R. A. FISHER, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, Vol. 7 (1936), pp. 179-188.
- [5] H. HOTELLING, "The Generalization of Student's Ratio," *Annals of Math. Stat.*, Vol. 2 (1931), pp. 359-378.
- [6] P. L. HSU, "Notes on Hotelling's Generalized  $T$ ," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 231-243.
- [7] S. KOŁODZIECZYK, "On an Important Class of Statistical Hypotheses," *Biometrika*, Vol. 27 (1935), pp. 161-190.
- [8] J. NEYMAN AND E. S. PEARSON, "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference," *Biometrika*, Vol. 20A (1928), pp. 175-240.
- [9] J. NEYMAN AND E. S. PEARSON, "The Testing of Statistical Hypotheses in Relation to Probabilities a Priori," *Proc. Camb. Phil. Soc.*, Vol. 29 (1933), p. 493.
- [10] J. NEYMAN AND E. S. PEARSON, "Contributions to the Theory of Testing Statistical Hypotheses," *Stat. Res. Mem.*, Vol. 1 (1936), pp. 1-37, Vol. 2 (1938), pp. 25-57.
- [11] P. C. TANG, "The Power Function of the Analysis of Variance Tests with Tables and Illustrations of Their Use," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 126-149 + viii.

- [12] S. S. WILKS, "Certain Generalizations in the Analysis of Variance," *Biometrika*, Vol. 24 (1932), pp. 471-494.
- [13] S. S. WILKS, "Moment-generating Operators for Determinants of Product Moments in Samples from a Normal System," *Ann. of Math.*, Vol. 35 (1934), pp. 312-340.
- [14] S. S. WILKS, "On the Independence of  $k$  Sets of Normally Distributed Statistical Variables," *Econometrica*, Vol. 3 (1935), pp. 309-326.
- [15] S. S. WILKS, "The Analysis of Variance for Two or More Variables," *Report of Third Annual Research Conference on Economics and Statistics*, (1937), pp. 82-85.
- [16] S. S. WILKS, "The Analysis of Variance and Covariance in Non-Orthogonal Data," *Metron*, Vol. XIII (1938), pp. 141-154.
- [17] J. WISHART, "The Generalized Product Moment Distribution in a Normal Distribution," *Biometrika*, Vol. 20<sup>A</sup> (1928), pp. 32-52.

PRINCETON UNIVERSITY,  
PRINCETON, N. J.

# STATISTICAL SEMINVARIANTS AND THEIR SETIMATES WITH PARTICULAR EMPHASIS ON THEIR RELATION TO ALGEBRAIC INVARIANTS

BY PAUL L. DRESSEL

## TABLE OF CONTENTS

	PAGE
Introduction.....	33
I. The Relation of the Algebraic Seminvariant Theory to the Moment Functions of Statistics	
1. Definitions.....	34
2. Complete Systems of Seminvariants.....	35
3. The MacMahon Non-Unitary Symmetric Function Principle.....	35
4. A Third Complete System of Seminvariants.....	36
5. Linearly Independent Seminvariants.....	36
6. The Roberts Theorem.....	37
7. Statistical Distributions Represented by Binary Forms.....	37
8. Three Systems of Statistical Seminvariants.....	38
9. Linearly Independent Statistical Seminvariants.....	39
10. Statistical Invariants.....	39
11. Seminvariants and Invariants of Samples.....	40
II. Estimates	
1. Power Product Seminvariants.....	41
2. Unbiased Estimates of Rational Integral Moment Functions.....	42
3. Computation of Estimates.....	42
4. The Dwyer Double Expansion Theorem.....	43
5. Estimates of all Seminvariants of Weight $\leq 8$ .....	44
6. Computation Checks.....	44
7. Estimates as Sums of Simple Seminvariants.....	50
8. The Estimates of the $\kappa$ 's.....	53
9. Other Simple Seminvariants which are Invariant under Estimate.....	54
10. Composite Seminvariants which are Invariant under Estimate.....	54
Conclusion.....	56
References.....	57

## INTRODUCTION

An important portion of algebraic invariant theory has been that devoted to a certain class of invariants called seminvariants, semi-invariants, or more rarely, half-invariants. Of these terms, "seminvariant" seems to be the one now commonly accepted. The same three terms have been applied at various times and by various writers to a system of moment functions of importance in statistical theory. The statistician using these terms has frequently done so with an apology for appropriating a term of the algebraist. As a portion of this paper we shall show that the moment functions of this system are actually algebraic seminvariants, and that there are other systems of moment functions which are equally entitled to the name seminvariant.



The study of the statistical seminvariants of a population leads naturally to consideration of the problem of obtaining from a sample unbiased estimates of the value of these seminvariants. Estimates of this kind have been defined and computed by previous authors, but no simple method of obtaining the estimates has been given. In this paper a simple procedure for calculation is given and it is furthermore demonstrated that these estimates form an important phase of statistical seminvariant theory.

The system of notation used for moment functions is that of R. A. Fisher, although the actual letters used in representing particular moment functions are not altogether the same as those used by Fisher. In general, a moment function of the population has been indicated by a Greek letter, the corresponding sample moment function by the corresponding English letter and the estimate by the corresponding capital English letter.

A list of references appears at the end of the paper. Each reference has been assigned a number and this number placed in square brackets is used in the body of the paper to indicate the reference. Pages of the reference are indicated by additional numbers inserted in the parentheses and separated from the reference number by a semicolon.

#### 1. THE RELATION OF THE ALGEBRAIC SEMINVARIANT THEORY TO THE MOMENT FUNCTIONS OF STATISTICS

The purposes of this chapter are: (1) to review briefly and give adequate references to certain important phases of algebraic seminvariant theory, (2) to apply this material to the moment functions of statistics.

1. **Definitions.** Any function of the coefficients of the binary form

$$(1) \quad f = \sum_{i=0}^n \binom{n}{i} a_i X^{n-i} Y^i, \quad a_0 \neq 0,$$

which is invariant under the transformation

$$(2) \quad X = \gamma_1 \xi + \gamma_2 \eta, \quad Y = \delta_1 \xi + \delta_2 \eta, \quad \Delta = \begin{vmatrix} \gamma_1 & \gamma_2 \\ \delta_1 & \delta_2 \end{vmatrix} \neq 0,$$

is called an invariant of the form  $f$ . See Dickson [1; 31-36].

Any function of the coefficients of  $f$  which is invariant under the transformation

$$(3) \quad X = \xi + \gamma \eta, \quad Y = \eta,$$

is called a seminvariant of  $f$ .

The two operators

$$(4) \quad \Omega = \sum_{i=1}^n i a_{i-1} \frac{\partial}{\partial a_i}, \quad \circ = \sum_{i=1}^n (n-i+1) a_i \frac{\partial}{\partial a_{i-1}},$$

are of fundamental importance in the theory of algebraic invariants and seminvariants and, indeed, invariants and seminvariants may be defined by means

of these operators. A necessary and sufficient condition that an homogeneous isobaric function of the coefficients of  $f$  be an invariant is that it be annihilated by both  $\Omega$  and  $\mathcal{O}$ . See Elliott [2; 113, 124]. The necessary and sufficient condition that an homogeneous isobaric function of the coefficients of  $f$  be a seminvariant is that it be annihilated by  $\Omega$ . See Elliott [2; 127].

It should be noted that there is nothing in the definitions above which requires that invariants or seminvariants be integral, although usually only this type is discussed. In what follows we shall find it more profitable to discuss homogeneous isobaric fractional seminvariants, the fractional quality resulting from the appearance of  $a_0$  in the denominator.

**2. Complete Systems of Seminvariants.** By direct application of the transformation (3) to  $f$  the system of seminvariants [1; 47]

$$(5) \quad A_r = \sum_{i=0}^r \binom{r}{i} \left( -\frac{a_1}{a_0} \right)^i \frac{a_{r-i}}{a_0}, \quad r \leq n,$$

is obtained. This system is a complete system, [2; 44, 205, 206], in the sense that all other seminvariants fractional in  $a_0$  and of degree 0 are expressible rationally and integrally in terms of this system.

Other such systems can be defined. The system of minimum degree seminvariants, the seminvariants of even weight being of degree 2 and those of odd weight being of degree 3, has played an important role in the algebraic seminvariant theory. Elliott [2; 207-209] discusses this system and gives the general formula for the even weight seminvariants of the system. So far as the present writer has been able to discover the general formula for the odd weight seminvariants has never been published, although Hammond [3] may have obtained it. After some lengthy but not difficult computation the result has been obtained, so that the last mentioned system of seminvariants is completely defined by

$$(6) \quad C_{2r+1} = \sum_{i=0}^r (-1)^{i+r} \binom{2r}{i+r} \frac{2i+1}{i+r+1} \frac{a_{r-i} a_{r+i+1}}{a_0^2} \\ + \sum_{i=0}^{2r} (-1)^{i+1} \binom{2r}{i} \frac{a_1 a_i a_{2r-i}}{a_0^3}.$$

It is easily demonstrated that for each of the above seminvariants, and in fact for any seminvariant, the sum of the numerical coefficients is zero. Dickson [1; 55] gives a suggestion leading to a very simple proof.

**3. The MacMahon Non-Unitary Symmetric Function Principle.** Denoting the roots of  $\sum_{i=0}^n \binom{n}{i} a_i X^{n-i} = 0$  by  $\alpha_1, \alpha_2, \dots, \alpha_n$ , the  $r$ -th power sum of these roots is defined by

$$(7) \quad s_r = \sum_{i=1}^n \alpha_i^r.$$

The form  $f$  may be written  $\prod_{i=1}^n (X - \alpha_i Y)$ .

By a result due to MacMahon [4; 131] the seminvariants of the form  $f$  are identical, except for numerical factors, with those symmetric functions of the roots of

$$(8) \quad g = \sum_{i=0}^n \frac{a_i}{i!} X^{n-i} = 0$$

which when expressed in terms of sums of powers of these roots do not contain  $s_1$ . MacMahon called such symmetric functions "non-unitary."

As a result of this theorem, MacMahon was able to discuss the seminvariants of a binary form of infinite order by discussing the non-unitary symmetric functions of the roots of  $\sum_{i=0}^{\infty} \frac{a_i}{i!} Y^i = 0$ .

**4. A Third Complete System of Seminvariants.** By application of the result stated in the previous section, a third complete system of seminvariants can be immediately obtained. Obviously the power sums  $s_r$ ,  $r > 1$ , are independent of  $s_1$ . By the Waring formula, Burnside and Panton [5; 91-92], if

$$\sum_{i=0}^{\infty} c_i Y^i = c_0 \prod_{i=0}^{\infty} (1 - \alpha_i Y)$$

then

$$(9) \quad s_r = \Sigma \frac{(-1)^{\rho} r(\rho-1)!}{\pi_1! \pi_2! \dots \pi_n!} \left(\frac{c_1}{c_0}\right)^{\pi_1} \left(\frac{c_2}{c_0}\right)^{\pi_2} \dots \left(\frac{c_n}{c_0}\right)^{\pi_n},$$

wherein

$$\rho = \sum_{i=1}^n \pi_i, \quad r = \sum_{i=1}^n i \pi_i.$$

Then for

$$(10) \quad g = \sum_{i=0}^{\infty} \frac{a_i}{i!} X^{n-i},$$

$$(r-1)! s_r = \Sigma \frac{(-1)^{\rho-1} r(\rho-1)! \left(\frac{a_1}{a_0}\right)^{\pi_1} \left(\frac{a_2}{a_0}\right)^{\pi_2} \dots \left(\frac{a_n}{a_0}\right)^{\pi_n}}{\pi_1! \pi_2! \dots \pi_n! (2!)^{\pi_2} \dots (n!)^{\pi_n}}.$$

Placing  $B_r = -(r-1)! s_r$ , the  $B$ 's form a complete system of seminvariants. This result has some interesting statistical connections which will be mentioned later.

**5. Linearly Independent Seminvariants.** It follows from the MacMahon non-unitary symmetric function principle, or it can be proved easily in other ways, that the number of linearly independent seminvariants of a given weight  $r$  is

equal to the number of partitions of  $r$  which contain no unit part. Furthermore we have at our disposal a simple method for obtaining a set of linearly independent seminvariants of any given weight.

For many purposes the power product defined by Dwyer [6; 13] is more useful than the customary monomial symmetric function. The power product is defined by the right hand member and indicated by the left hand member of

$$(11) \quad (q_1 \dots q_r) = \sum_{i_1 \neq i_2 \neq \dots \neq i_r} \alpha_{i_1}^{q_1} \alpha_{i_2}^{q_2} \dots \alpha_{i_r}^{q_r},$$

where, for convenience,  $q_1 \geq q_2 \geq \dots \geq q_r$ . The monomial symmetric function which will be denoted by  $M(q_1 \dots q_r)$  is related to the power product by the identity

$$(12) \quad \pi_1! \dots \pi_r! M(q_1^{r_1} q_2^{r_2} \dots q_r^{r_r}) = (q_1^{r_1} q_2^{r_2} \dots q_r^{r_r}),$$

so that a distinction occurs only when there are repeated exponents in the summation of (11).

If we desire a system of linearly independent seminvariants of weight 6, by the MacMahon principle we need only to compute the values of the power products (6), (42), (33), (222) in terms of the  $a$ 's. In a somewhat different form these will be presented later.

**6. The Roberts Theorem.** Roberts, see [2; 231] and [5; 108], demonstrated the existence of a duality relationship between power sums,  $s$ 's, and coefficients,  $a$ 's such that corresponding to any seminvariant in terms of  $a$ 's there exists a seminvariant in terms of  $s$ 's obtained by replacing  $a_i$  by  $s_i$ . The proof consists of showing that the annihilator for seminvariants in terms of power sums is identical in form with  $\Omega$ ,  $a_i$  being replaced by  $s_i$ .

As a result of this duality, each of the systems of seminvariants which have been obtained yields, upon replacement of  $a_i$  by  $s_i$ , another system of seminvariants. In particular cases it may happen that the systems are identical when the identities connecting the  $a_i$  and  $s_i$  are taken into consideration.

We next wish to show that the systems of power sum seminvariants thus obtained either are identical with certain well known statistical moment functions or lead to new ones.

**7. Statistical Distributions Represented by Binary Forms.** The fact that statistical distributions may be represented by polynomials has long been recognized by statisticians, see Thiele [7; 24-26] and Bertelsen [8]. Indeed it was this fact which led Thiele to the definition of the seminvariants now called by his name. If we have given  $n$  observations  $\alpha_1, \alpha_2, \dots, \alpha_n$ , form the polynomial.

$$(13) \quad F = \prod_{i=1}^n (X - \alpha_i) = \sum_{i=0}^n \binom{n}{i} \frac{a_i}{a_0} X^{n-i}.$$

$F$  is not a binary form, but the seminvariant theory of binary forms is applicable since seminvariants are functions of the differences of the roots and are independent of the  $X$  and  $Y$ , which appear merely as convenient symbols to indicate the various terms of the algebraic form.

For distributions containing an infinite number of items the form  $F$  is of infinite order, but discussion of its seminvariants may be carried on by use of the MacMahon principle given in section 3.

**8. Three Systems of Statistical Seminvariants.** Before exhibiting some systems of statistical seminvariants it may be well to consider the meaning of "statistical seminvariant," for this phrase has been undefined. In fact the use of the phrase is merely a matter of convenience in that it emphasizes the fact that seminvariant moment functions have not previously been regarded as algebraic seminvariants. As used here a statistical seminvariant is an algebraic seminvariant which has some application in statistical theory.

The system of seminvariants (5) yields by application of the Roberts' Theorem the well known system of statistical seminvariants usually called central moments. If  $\mu'_r = \frac{s_r}{n} = \frac{s_r}{s_0}$ , the general formula may be written

$$(14) \quad \mu_r = \sum_{i=0}^r \binom{r}{i} \mu'_{r-i} (-\mu'_1)^i.$$

The system of seminvariants (6) likewise leads to

$$(15) \quad \kappa_{2r} = \frac{1}{2} \sum_{i=0}^{2r} (-1)^i \binom{2r}{i} \mu'_i \mu'_{2r-i},$$

$$\kappa_{2r+1} = \sum_{i=0}^r (-1)^{i+r} \binom{2r}{i+r} \frac{2i+1}{i+r+1} \mu'_{r-i} \mu'_{r+i+1} + \sum_{i=0}^{2r} (-1)^{i+1} \binom{2r}{i} \mu'_i \mu'_i \mu'_{2r-i},$$

a system which seems never to have been used by statisticians.

The system (10) leads to the well known Thiele seminvariants

$$(16) \quad \lambda_r = \sum \frac{(-1)^{r-1} r! (\rho-1)! (\mu'_1)^{\pi_1} (\mu'_2)^{\pi_2} \dots (\mu'_r)^{\pi_r}}{\pi_1! \pi_2! \dots \pi_r! (2!)^{\pi_2} \dots (r!)^{\pi_r}}.$$

From sections 3 and 4 it is apparent that the general formula for the Thiele seminvariants is a special case of the Waring formula for power sums in terms of coefficients. It does not seem that this fact has been previously recognized. An equivalent way of stating this idea is to say that the Thiele seminvariant  $\lambda_r$  is, except for the factor  $-(r-1)!$ , the sum of the  $r$ -th powers of the roots of the equations obtained by setting the moment generating function,

$$M_\pi(Y) = \sum_{i=0}^{\infty} \mu'_i \frac{Y^i}{i!},$$

equal to zero.

It is of historical interest to note that MacMahon published his non-unitary function principle and the resulting set of seminvariants in 1884. Cayley [8] published an article in 1885 dealing with this same system. Roberts' Theorem having been known for some time (probably about 20 years), it seems probable that MacMahon and Cayley were aware of the Thiele seminvariants four to five years before Thiele's definition [9] by an entirely different method.

**9. Linearly Independent Statistical Seminvariants.** At the end of section 5 a method was indicated whereby a complete set of linearly independent seminvariants of a given weight  $r$  could be obtained. It has been noted previously that the one part symmetric function  $s_r$  or  $(r)$  leads to the Thiele seminvariant  $\lambda_r$ . As a further illustration consider the power product (22). From a table of symmetric functions we find that

$$\begin{aligned}(22) &= \frac{2a_4}{4!a_0} - \frac{2a_3a_1}{3!a_0^2} + \frac{a_2^2}{2!2!a_0^2} \\ &= \frac{2}{4!} \left( \frac{a_4}{a_0} - \frac{4a_3a_1}{a_0^2} + \frac{3a_2^2}{a_0^2} \right),\end{aligned}$$

and by the Roberts' Theorem the statistical seminvariant

$$\frac{2}{4!} (\mu'_4 - 4\mu'_3\mu'_1 + 3\mu_2'^2)$$

is obtained. In similar fashion a system of linearly independent seminvariants of weight  $\leq 8$  have been computed and are given in Table I. For the sake of brevity they are expressed in terms of central moments. Hence the degree, by which is meant the maximum degree in the  $\mu''$ 's, is not apparent in the table. This definition of degree associates with the statistical seminvariant the degree (in the usual sense) of the corresponding homogeneous integral seminvariant.

**10. Statistical Invariants.** If the transformation

$$(17) \quad x = \xi + mk\eta, \quad y = m\eta$$

is applied to the binary form  $f$  and, if, in particular

$$k = -\frac{a_1}{a_0} \quad \text{and} \quad m = \left[ \frac{a_2}{a_0} - \frac{a_1^2}{a_0^2} \right]^{-1}$$

one system of invariants of  $f$  under this transformation is found to be

$$(18) \quad D_r = A_r/A_2^{\frac{1}{2}r}, \quad r \leq n,$$

where  $A_r$  is defined in (5). By the Roberts Theorem we obtain the fact that the standard moment  $\mu_r/\mu_2^{\frac{1}{2}r}$  is an invariant of  $f$  under this transformation. Thus the standard moments, or standard seminvariants in general, have also an algebraic connection. The effect of the transformation (17) on the roots of  $f$  is indicated by

$$x - \alpha_i y = \xi + mk\eta - m\alpha_i\eta = \xi - m(\alpha_i - k)\eta.$$

If  $m$  and  $k$  are defined as above, the result is the equivalent of measuring in standard units denoted by  $\frac{\alpha_i - \mu_1'}{\sqrt{\mu_2}}$ .

The system (18) is not a system of algebraic invariants, for algebraic invariants must be invariant under rotation, translation and change of scale, or stretching. The component parts of the above system are invariant only under the last two

TABLE I  
*Linearly Independent Seminvariants of Weight  $\leq 8$*

Weight	Degree	Seminvariants	Weight	Degree	Seminvariants
6	6	$\mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 30\mu_2^3$	0	1	$\mu_0$
	4	$\mu_6 + 5\mu_4\mu_2 - 10\mu_3^2 - 30\mu_2^3$	2	2	$\mu_2$
	3	$\mu_6 - 15\mu_4\mu_2 + 20\mu_3^2 + 30\mu_2^3$	3	3	$\mu_3$
	2	$\mu_6 + 15\mu_4\mu_2 - 10\mu_3^2$	4	4	$\mu_4 - 3\mu_2^2$
7	7	$\mu_7 - 21\mu_5\mu_2 - 35\mu_4\mu_3 + 56\mu_3\mu_2^2$		2	$\mu_4 + 3\mu_2^2$
	5	$\mu_7 + 9\mu_5\mu_2 - 35\mu_4\mu_3 - 90\mu_3\mu_2^2$	5	5	$\mu_5 - 10\mu_3\mu_2$
	4	$\mu_7 - 21\mu_5\mu_2 + 25\mu_4\mu_3 + 30\mu_3\mu_2^2$		3	$\mu_5 + 2\mu_3\mu_2$
	3	$\mu_7 + 9\mu_5\mu_2 - 5\mu_4\mu_3$			
8	8	$\mu_8 - 28\mu_6\mu_2 - 56\mu_5\mu_3 - 70\mu_4^2 + 210\mu_4\mu_2^2 + 280\mu_3^2\mu_2 - 105\mu_2^4$			
	6	$\mu_8 + 14\mu_6\mu_2 - 56\mu_5\mu_3 - 35\mu_4^2 - 210\mu_4\mu_2^2 + 140\mu_3^2\mu_2 + 630\mu_2^4$			
	5	$\mu_8 - 28\mu_6\mu_2 + 49\mu_5\mu_3 - 35\mu_4^2 + 420\mu_4\mu_2^2 - 490\mu_3^2\mu_2 - 630\mu_2^4$			
	4	$\mu_8 - 28\mu_6\mu_2 - 56\mu_5\mu_3 + 105\mu_4^2 - 420\mu_4\mu_2^2 + 560\mu_3^2\mu_2 + 630\mu_2^4$			
	4	$\mu_8 + 14\mu_6\mu_2 - 56\mu_5\mu_3 + 35\mu_4^2 - 210\mu_4\mu_2^2 + 140\mu_3^2\mu_2$			
	3	$\mu_8 - 7\mu_6\mu_2 + 49\mu_5\mu_3 - 35\mu_4^2 + 105\mu_4\mu_2^2 - 70\mu_3^2\mu_2$			
	2	$\mu_8 + 28\mu_6\mu_2 - 56\mu_5\mu_3 + 35\mu_4^2$			

types of transformation. In statistics translation and change of scale ordinarily constitute the only desired transformations so that the standard seminvariants

$\frac{\mu_r}{\mu_2^{\frac{1}{2}r}}, \frac{\lambda_r}{\lambda_2^{\frac{1}{2}r}}, \frac{\kappa_r}{\kappa_2^{\frac{1}{2}r}}, \dots$  might well be called statistical invariants.

**11. Seminvariants and Invariants of Samples.** Consideration of the definition of seminvariants and invariants shows that:

1. A seminvariant is a seminvariant not because it is a function of deviations from the mean, but because it is a function of the differences of the observations;

2. An invariant is an invariant not because it is a seminvariant divided by the standard deviation raised to the proper power, but because it is a ratio of two seminvariants which are of the same order in powers of the observations.

These facts are important from the statistics viewpoint because they show that seminvariants and invariants of samples are also seminvariants and invariants of the population from which the samples are drawn.

## II. ESTIMATES

**1. Power Product Seminvariants.** The Roberts Theorem set up a duality relationship between seminvariants expressed in terms of coefficients and seminvariants in terms of power sums. It can be shown that corresponding to each pair thus determined there exists a third seminvariant expressed in terms of power products. This leads to what may be called a triple system of seminvariants, the interrelationships being most apparent when all three seminvariants are expressed in terms of the notation defined by (11). The seminvariant  $a_3 - \frac{3a_2a_1}{a_0} + \frac{2a_1^3}{a_0^3}$  becomes in this notation

$$(111) - \frac{3(11)(1)}{n^{(2)}n} + \frac{2(1)^3}{n^3}$$

The corresponding power sum seminvariant is

$$(3) - \frac{3(2)(1)}{n^2} + \frac{2(1)^3}{n^3},$$

while the power product seminvariant just mentioned is

$$(3) - \frac{3(21)}{n^{(2)}} + \frac{2(111)}{n^{(3)}}.$$

The value of the power product notation lies in the fact that the numerical coefficients of the three seminvariants are then identical, while this is not the case when monomial and elementary symmetric functions are used.

Perhaps a few remarks are in order in regard to the proof of the relationship above expressed. The annihilator, corresponding to  $\Omega$ , for seminvariants in terms of roots is, see [2; 230-31],

$$-D = \sum_{i=1}^n \frac{\partial}{\partial \alpha_i}.$$

It is easy to see that

$$-D \left[ \frac{(p_1^{\tau_1} p_2^{\tau_2} \dots p_s^{\tau_s})}{n^{(\rho)}} \right] = \frac{1}{n^{(\rho)}} \sum_{i=1}^s \pi_i p_i (p_1^{\tau_1} p_2^{\tau_2} \dots p_i^{\tau_i-1}, p_i - 1, \dots p_s^{\tau_s}),$$

and also that,

$$\frac{(p_1^{\tau_1} p_2^{\tau_2} \dots p_{s-1}^{\tau_{s-1}})}{n^{(\rho)}} = \frac{(n - \rho + 1)(p_1^{\tau_1} \dots p_{s-1}^{\tau_{s-1}})}{n^{(\rho)}} = \frac{(p_1^{\tau_1} \dots p_{s-1}^{\tau_{s-1}})}{n^{(\rho-1)}}.$$

Since

$$\Omega \left[ \frac{(p_1)^{\tau_1} (p_2)^{\tau_2} \dots (p_s)^{\tau_s}}{n^{\rho}} \right] = \frac{1}{n^{\rho}} \sum_{i=1}^s \pi_i p_i (p_1)^{\tau_1} (p_2)^{\tau_2} \dots (p_i)^{\tau_i-1} (p_i - 1) \dots (p_s)^{\tau_s},$$



and

$$\frac{(p_1)^{\tau_1}(p_2)^{\tau_2} \dots (p_{s-1})^{\tau_{s-1}}(0)}{n^s} = \frac{n(p_1)^{\tau_1}(p_2)^{\tau_2} \dots (p_{s-1})^{\tau_{s-1}}}{n^s} = \frac{(p_1)^{\tau_1} \dots (p_{s-1})^{\tau_{s-1}}}{n^{s-1}},$$

it becomes evident that corresponding to any power sum seminvariant there exists a power product seminvariant with the same numerical coefficients. The converse is also true.

**2. Unbiased Estimates of Rational Integral Moment Functions.** If  $\tau$  represents a population parameter, and if  $t$  represents such a function of  $n$  observations that the expected value of  $t$  is equal to  $\tau$ ; then  $t$  is said to be an unbiased estimate of  $\tau$ . See Tschuprow [11; 74-75], Bertilsen [8; 144], and Fisher [12].

Let  $(p_1 p_2 \dots p_s)$  denote a power product computed from a sample, the sample being from an infinite population. Then it is well known that

$$E \left[ \frac{(p_1 p_2 \dots p_s)}{n^{(s)}} \right] = \mu'_{p_1} \mu'_{p_2} \dots \mu'_{p_s},$$

$n$  being the number of items in the sample. If  $E^{-1}$  be interpreted as "unbiased estimate of," the above relation may also be written

$$(19) \quad E^{-1}[\mu'_{p_1} \mu'_{p_2} \dots \mu'_{p_s}] = \frac{(p_1 p_2 \dots p_s)}{n^{(s)}},$$

and it is seen at once that the power product seminvariants defined in section 1, if computed from a sample of  $n$  observations, are the unbiased estimates of the corresponding power sum seminvariants of the infinite population from which the sample is drawn.

This provides an algebraic interpretation as well as a different approach to a topic which has already aroused considerable interest among statisticians. In 1927 Bertilsen [8; 144] gave the estimates of the first four Thiele seminvariants of the population in terms of Thiele seminvariants of the sample. In 1929 R. A. Fisher [12] also obtained these results and gave in addition the estimates of the fifth and sixth Thiele seminvariants. His results are in terms of sample moments. In 1937, P. S. Dwyer [13; 26] gave the estimates of the first five population central moments and indicated also means for obtaining the estimate of any rational integral isobaric moment function.

In the remainder of this chapter

- (1) Dwyer's method will be extended and perhaps somewhat simplified,
- (2) certain properties of this type of estimate will be pointed out,
- (3) estimates of all seminvariants of weight  $\leq 8$  will be made available.

**3. Computation of Estimates.** From the relationship (19) it is possible to write down immediately in a simple, although not immediately useful, form the estimate of any rational integral moment function. Thus the fourth Thiele seminvariant  $\lambda_4$  is given by

$$\lambda_4 = \mu'_4 - 4\mu'_3\mu'_1 - 3\mu'^2_2 + 12\mu'_2\mu'^2_1 - 6\mu'^4_1,$$

so that the estimate of  $\lambda_4$  is

$$L_4 = \frac{(4)}{n} - \frac{4(31)}{n^{(2)}} - \frac{3(22)}{n^{(2)}} + \frac{12(211)}{n^{(3)}} - \frac{6(1111)}{n^{(4)}}.$$

Since power products are difficult to compute directly, it is necessary to express the estimates in terms of power sums. Dwyer [6; 30-33] gave a complete discussion of the problem of expanding power products in terms of power sums and also gave tables of power products in terms of power sums for weights  $\leq 6$ . By use of (12) it is also possible to use tables giving monomial symmetric functions in terms of power sums. One table by J. R. Roe [14; plate 18] includes all cases of weight  $\leq 10$ .

By use of such a table we find

$$(31) = -(4) + (3)(1),$$

$$(22) = -(4) + (2)(2),$$

$$(211) = 2(4) - 2(3)(1) - (2)(2) + (2)(1)^2,$$

$$(1111) = -6(4) + 8(3)(1) + 3(2)(2) - 6(2)(1)^2 + (1)^4.$$

If these results are substituted in  $L_4$  above and like terms are collected, it is found that

$$n^{(4)}L_4 = n^2(n+1)(4) - 4n(n+1)(3)(1) - 3n(n-1)(2)^2 + 12n(2)(1)^2 - 6(1)^4,$$

a result which agrees with that given by R. A. Fisher [12].

**4. The Dwyer Double Expansion Theorem.** The Dwyer double expansion theorem, [6; 34] and [11; 37-39], states that if any isobaric sum of power products of weight  $r$  indicated by

$$(20) \quad \Sigma \frac{r!}{(q_1!)^{r_1} \dots (q_t!)^{r_t} \pi_1! \dots \pi_t!} b_{q_1^{r_1} \dots q_t^{r_t}} (q_1^{r_1} \dots q_t^{r_t})$$

be expanded in terms of power sums in a form indicated by

$$(21) \quad \Sigma \frac{r!}{(p_1!)^{r_1} \dots (p_s!)^{r_s} \pi_1! \dots \pi_s!} a_{p_1^{r_1} \dots p_s^{r_s}} (p_1)^{r_1} \dots (p_s)^{r_s},$$

then the coefficient  $a_r$  of the power sum  $(r)$  is given by

$$(22) \quad a_r = \Sigma (-1)^{r-1} \frac{(\rho-1)! r!}{(p_1!)^{r_1} \dots (p_s!)^{r_s} \pi_1! \dots \pi_s!} b_{p_1^{r_1} \dots p_s^{r_s}},$$

and that the coefficient  $a_{r_1 \dots r_m}$  of  $(r_1)(r_2) \dots (r_m)$  is

$$(23) \quad a_{r_1 \dots r_m} = a_{r_1} a_{r_2} \dots a_{r_m}.$$

The barred product indicates a symbolic multiplication by suffixing of subscripts which is exemplified by

$$a_2 a_2 = (b_2 - 3b_{21} + 2b_{111})(b_2 - b_{11}) = b_{22} - b_{211} - 3b_{221} + 5b_{2111} - 2b_{11111} = a_{22}.$$

The application of this theorem to the present problem eliminates the use of tables and permits the independent computation of the coefficient of any particular products of power sums in the expansion in terms of power sums of any given estimate. The illustration given by Dwyer [13; 39, 40] exemplifies both of these points very well.

**5. Estimates of all Seminvariants of Weight  $\leq 8$ .** If the estimates of any complete system of seminvariants and all products of these seminvariants up to and including weight  $r$  are known, then the estimates of all seminvariants of weight  $\leq r$  are obtainable as a linear combination of these known estimates. For example, suppose that we know the estimates of all Thiele seminvariants of weight  $\leq 5$  and wish to find the estimate of  $\mu_5$ . Since  $\mu_5 = \lambda_5 + 10\lambda_3\lambda_2$ ,

$$E^{-1}[\mu_5] = M_5 = E^{-1}[\lambda_5] + 10E^{-1}[\lambda_3\lambda_2] = L_5 + 10L_{32}.$$

In table II are given the estimates of all Thiele seminvariants and all products of Thiele seminvariants of weight  $\leq 8$ . From this table the expressions for  $L_5$  and  $L_{32}$  are obtained and, by taking the combination indicated above, it is seen that

$$\begin{aligned} n^{(5)}M_5 &= (n^4 - 5n^3 + 10n^2)(5) - 5(n^3 - 5n^2 + 10n)(4)(1) \\ &\quad - 10(n^2 - n)(3)(2) + 10(n^2 - 4n + 8)(3)(1)^2 \\ &\quad + 30(n - 2)(2)^2(1) - 10n(2)(1)^3 + 4(1)^5, \end{aligned}$$

a result which checks with that given by Dwyer [13; 27]. In similar fashion the estimate of any other seminvariant of weight  $\leq 8$  can be obtained by use of table II.

**6. Computation Checks.** There are a number of checks which can be applied to the entries in table II. These may be of interest simply as properties of the estimates, and they may be of use in correcting errors which may possibly have crept into the tables.

When any power product of more than one part is expanded into power sums, the sum of the numerical coefficients of the expansion is zero. To prove this we need only to consider a set of observations of which one observation is unity and the rest are all zero. Then any power product of two or more parts is necessarily zero and all power sums are equal to unity. Hence the initial statement of the paragraph follows immediately.

From this fact it is apparent that the sum of the coefficients of  $L_r$  is  $\frac{1}{n}$ , and the sum of the coefficients of  $L_{r_1 r_2 \dots r_s}$  is zero. Thus for  $L_4$  we have  $\frac{n^3 + n^2 - 4(n^2 + n) - 3(n^2 - n) + 12n - 6}{n^{(4)}} = \frac{1}{n}$ , and for  $L_{22}$  the sum of the coefficients is

$$\frac{1}{n^{(4)}} [-n^3 + n + 4n - 4 + n^2 - 3n + 3 - 2n + 1] = 0.$$

TABLE II  
Estimates of All Thiele Seminvariants and Their Products of Weight  $\leq 8$

$w = 1$		$w = 3$		$w = 4$		$n^{(1)}L_2$		$w = 5$		$n^{(1)}L_4$		$n^{(1)}L_6$	
(1)	1	(3)	$n^3$	(4)	$n^3 + n^2$	(4)	$-(n^2 - n)$	(5)			$n^4 + 5n^2$		$-(n^3 - n^2)$
		(2)(1)	$-3n$	(3)(1)	$-4(n^2 + n)$	(3)	$4(n - 1)$	(4)(1)			$-5(n^3 + 5n^2)$		$5(n^2 - n)$
		(1) <sup>2</sup>	2	(2) <sup>2</sup>	$-3(n^2 - n)$	(2)	$n^2 - 3n + 3$	(3)(2)			$-10(n^3 - n^2)$		$n^3 - 2n^2 + 2n$
				(2)(1) <sup>2</sup>	$12n$		$-2n$	(3)(1) <sup>2</sup>			$20(n^2 + 2n)$		$-(n^2 + 8n - 8)$
				(1) <sup>4</sup>	$-6$		1	(2) <sup>2</sup> (1)			$30(n^2 - n)$		$-3(n^3 - 2n + 2)$
								(2)(1) <sup>3</sup>			$-60n$		$+5n$
								(1) <sup>5</sup>			$24$		$-2$

$w = 2$		$w = 6$	
(2)	$n$	(6)	
(1) <sup>2</sup>	$-1$	(5)(1)	

$n^{(1)}L_2$		$n^{(1)}L_4$		$n^{(1)}L_6$		$n^{(1)}L_8$	
	$n^3 + 16n^4 + 11n^2 - 4n^2$		$-(n^4 + 2n^2 - 7n^2 + 4n)$		$-(n^4 - 2n^2 + 5n^2 - 4n)$		$2(n^2 - 3n^2 + 2n)$
	$-6(n^4 + 16n^3 + 11n^2 - 4n)$		$6(n^2 + 2n^2 - 7n + 4)$		$6(n^2 - 2n^2 + 5n - 4)$		$-12(n^2 - 3n + 2)$
	$-15(n^4 - 4n^2 - n^2 + 4n)$		$n^4 - 10n^2 + 45n - 60$		$3(2n^2 - 5n^2 - 5n + 20)$		$-3(n^2 - 7n^2 + 20n - 20)$
	$-10(n^4 - 2n^2 + 5n^2 - 4n)$		$2(2n^2 - 5n^2 - 5n + 20)$		$n^4 - 8n^2 + 25n^2 - 10n - 40$		$-2(3n^2 - 15n + 20)$
	$30(n^2 + 9n^2 + 2n)$		$-(n^2 + 15n^2 + 20n - 60)$		$-3(7n^2 - 15n + 20)$		$3(n^2 + 3n - 10)$
	$120(n^3 - n)$		$-4(n^2 + 6n^2 + 25n + 30)$		$-6(n^2 - 4n^2 + 15n - 20)$		$12(n^2 - 4n + 6)$
	$30(n^2 - 3n^2 + 2n)$		$-3(n^2 - 7n^2 + 20n - 20)$		$-3(3n^2 - 15n + 20)$		$n^2 - 9n^2 + 29n - 30$
	$-120(n^2 + 3n)$		$4(n^2 + 9n - 10)$		$4(n^2 + 3n)$		$-4(3n - 5)$
	$-270(n^2 - n)$		$3(5n^2 - 9n + 10)$		$9(n^2 - n)$		$-3(n^2 - 3n + 5)$
	$360n$		$-18n$		$-12n$		$3n$
	$-120$		6		4		$-1$

TABLE II—Continued

$w = 7$	$n^{(v)}L_l$	$n^{(v)}L_{eq}$	$n^{(v)}L_{us}$	$n^{(v)}L_{ss}$
(7)	$n^6 + 42n^5 + 119n^4 - 42n^3$	$-(n^5 + 12n^4 - 31n^3 + 18n^2)$	$-(n^5 + 5n^3 - 6n^2)$	$2(n^4 - 3n^3 + 2n^2)$
(6)(1)	$-7(n^5 + 42n^4 + 119n^3 - 42n^2)$	$7(n^4 + 12n^3 - 31n^2 + 18n)$	$7(n^4 + 5n^2 - 6n)$	$-14(n^3 - 3n^2 + 2n)$
(5)(2)	$-21(n^5 + 12n^4 - 31n^3 + 18n^2)$	$n^5 + 6n^4 - 28n^3 + 99n^2 - 162n$	$3(3n^4 - 10n^3 + 5n^2 + 18n)$	$-2(n^4 - 6n^3 + 17n^2 - 18n)$
(4)(3)	$-35(n^5 + 5n^3 - 6n^2)$	$5(3n^4 - 10n^3 + 5n^2 + 18n)$	$n^5 - 6n^4 + 30n^3 - 35n^2 - 30n$	$-(n^4 - n^3 - 10n^2 + 20n)$
(5)(1) <sup>2</sup>	$42(n^4 + 27n^3 + 44n^2 - 12n)$	$-(n^4 + 27n^3 + 224n^2 - 552n + 216)$	$-6(5n^3 - 5n^2 + 20n - 12)$	$2(n^3 + 15n^2 - 40n + 24)$
(4)(2)(1)	$210(n^4 + 6n^3 - 13n^2 + 6n)$	$-5(n^4 + 15n^3 - 58n^2 + 114n - 108)$	$-3(n^4 + 9n^3 - 20n^2 - 10n + 60)$	$13n^3 - 63n^2 + 140n - 120$
(3) <sup>2</sup> (1)	$140(n^4 + 5n^3 - 6n)$	$-20(3n^3 - 10n^2 + 5n + 18)$	$-4(n^4 - 6n^3 + 30n^2 - 35n - 30)$	$4(n^3 - n^2 - 10n + 20)$
(3)(2) <sup>2</sup>	$210(n^4 - 3n^3 + 2n^2)$	$-10(n^4 - 6n^3 + 17n^2 - 18n)$	$-3(n^4 - n^3 - 10n^2 + 20n)$	$n^4 - 7n^3 + 21n^2 - 20n$
(4)(1) <sup>3</sup>	$-210(n^3 + 13n^2 + 6n)$	$5(n^3 + 19n^2 + 36n - 108)$	$2(n^3 + 34n^2 - 45n + 90)$	$-4(3n^2 + 2n - 15)$
(3)(2)(1) <sup>2</sup>	$-1260(n^3 + n^2 - 2n)$	$30(n^3 + 7n^2 - 28n + 36)$	$12(2n^3 - 2n^2 + 15n - 30)$	$-2(n^3 + 12n^2 - 48n + 60)$
(2) <sup>2</sup> (1)	$-630(n^3 - 3n^2 + 2n)$	$30(n^3 - 6n^2 + 17n - 18)$	$9(n^3 - n^2 - 10n + 20)$	$-3(n^3 - 7n^2 + 21n - 20)$
(3)(1) <sup>4</sup>	$840(n^2 + 4n)$	$-20(n^2 + 10n - 12)$	$-14(n^2 + 4n)$	$n^3 + 24n - 40$
(2) <sup>3</sup> (1) <sup>2</sup>	$2520(n^2 - n)$	$-30(3n^2 - 5n + 6)$	$-42(n^2 - n)$	$2(4n^2 - 9n + 15)$
(2)(1) <sup>5</sup>	$-2520n$	$84n$	$42n$	$-7n$
$\Gamma(1)^7$	$720$	$-24$	$-12$	$2$

TABLE II—Continued

$w = 8$	$n^{(0)}L_0$	$n^{(0)}L_2$	$n^{(0)}L_3$
(6)	$n^3 + 99n^2 + 757n^4 + 141n^4 - 398n^3 + 120n^3$	$-(n^6 + 37n^5 - 39n^4 - 157n^3 + 278n^2 - 120n)$	$-(n^6 + 9n^5 - 23n^4 + 111n^3 - 218n^2 + 120n)$
(7)(1)	$-8(n^6 + 99n^5 + 757n^4 + 141n^4 - 398n^3 + 120n)$	$8(n^6 + 37n^4 - 39n^3 - 157n^2 + 278n - 120)$	$8(n^6 + 9n^4 - 23n^3 + 111n^2 - 218n + 120)$
(6)(2)	$-28(n^6 + 37n^5 - 39n^4 - 157n^3 + 278n^2 - 120n)$	$n^6 + 20n^5 + 3n^4 - 336n^3 + 1736n^2 - 4424n + 3360$	$13n^5 - 14n^4 - 57n^3 - 406n^2 + 2744n - 3360$
(6)(3)	$-58(n^6 + 9n^5 - 23n^4 + 111n^3 - 218n^2 + 120n)$	$2(13n^5 - 14n^4 - 57n^3 - 406n^2 + 2744n - 3360)$	$n^6 + n^5 - 51n^4 + 527n^3 - 1134n^2 - 2128n + 6720$
(4)*	$-35(n^6 + n^5 + 33n^4 - 121n^3 + 206n^2 - 120n)$	$5(3n^5 - 11n^4 + 11n^3 + 119n^2 - 602n + 840)$	$5(n^6 + n^4 - 39n^3 + 119n^2 + 182n - 840)$
(6)(1)*	$58(n^6 + 68n^4 + 359n^3 - 8n^2 - 60n)$	$-(n^6 + 48n^4 + 1039n^3 - 1428n^2 - 2660n + 3360)$	$-(41n^4 + 238n^3 - 701n^2 + 2702n - 3360)$
(6)(2)(1)	$336(n^6 + 23n^4 - 31n^3 - 23n^2 + 30n)$	$-6(n^6 + 33n^4 - 11n^3 - 393n^2 + 1330n - 1680)$	$-3(n^6 + 27n^4 - 79n^3 + 413n^2 - 1946n + 3360)$
(4)(3)(1)	$560(n^6 + 5n^4 + 5n^3 - 5n^2 - 6n)$	$-10(25n^4 - 58n^3 - 13n^2 + 70n + 336)$	$-5(n^6 + 9n^4 - 43n^3 + 215n^2 - 182n - 672)$
(4)(2)*	$420(n^6 + 2n^4 - 25n^3 + 46n^2 - 24n)$	$-15(n^6 - 2n^4 - 27n^3 + 236n^2 - 700n + 672)$	$-15(5n^4 - 26n^3 - 25n^2 + 406n - 672)$
(3)*(2)	$560(n^5 - 4n^4 + 11n^3 - 20n^2 + 12n)$	$-10(n^6 - 5n^5 - 96n^4 + 532n^3 - 672)$	$-10(n^6 - 11n^4 + 61n^3 - 101n^2 - 238n + 672)$
(5)(1)*	$-336(n^4 + 38n^3 + 99n^2 - 18n)$	$6(n^4 + 38n^3 + 339n^2 - 736n)$	$2(n^4 + 68n^3 + 159n^2 - 288n + 756)$
(4)(2)(1)*	$-2520(n^4 + 10n^3 - 17n^2 + 6n)$	$45(n^4 + 18n^3 - 41n^2 + 22n)$	$15(n^4 + 28n^3 - 113n^2 + 264n - 252)$
(3)*(1)*	$-1680(n^4 + 2n^2 + 7n^2 - 10n)$	$10(n^4 + 50n^3 - 121n^2 - 122n + 672)$	$20(n^4 - n^3 + 9n^2 + 57n - 210)$
(3)(2)*(1)	$-5040(n^4 - 2n^2 - n^2 + 2n)$	$120(n^4 - n^3 - 16n^2 + 70n - 84)$	$60(n^4 - 6n^3 + 35n^2 - 126n + 168)$
(2)*	$-630(n^4 - 6n^3 + 11n^2 - 6n)$	$30(n^4 - 12n^3 + 62n^2 - 147n + 126)$	$90(n^4 - 10n^3 + 35n^2 - 42)$
(4)(1)*	$1680(n^3 + 17n^2 + 12n)$	$-30(n^4 + 23n^3 + 54n^2 - 168)$	$-10(n^5 + 38n^4 - 45n^3 + 126)$
(3)(2)(1)*	$13440(n^3 + 2n^2 - 3n)$	$-240(n^3 + 8n^2 - 31n + 42)$	$-20(7n^3 + 2n^2 + 45n - 126)$
(2)*(1)*	$10080(n^3 - 3n^2 + 2n)$	$-60(5n^3 - 27n^2 + 76n - 84)$	$-90(n^3 - 2n^2 - 5n + 14)$
(3)(1)*	$-6720(n^3 + 5n)$	$120(n^3 + 11n - 14)$	$64(n^3 + 5n)$
(2)*(1)*	$-25200(n^2 - n)$	$90(7n^2 - 11n + 14)$	$240(n^2 - n)$
(2)(1)*	$20160n$	$-480n$	$-192n$
(1)*	$-5040$	$120$	$48$

TABLE II—Continued

$w = 8$	$n^{(6)}L_{44}$	$n^{(6)}L_{43}$	$n^{(6)}L_{33}$	$n^{(6)}L_{32}$
(8)	$-(n^6 + n^5 + 33n^4 - 121n^3 + 206n^2 - 120n)$	$2(n^5 + 2n^4 - 25n^3 + 46n^2 - 24n)$	$2(n^5 - 4n^4 + 11n^3 - 20n^2 + 12n)$	$-6(n^4 + 6n^3 + 11n^2 - 6n)$
(7)(1)	$8(n^5 + n^4 + 33n^3 - 121n^2 + 206n - 120)$	$-16(n^4 + 2n^3 - 25n^2 + 46n - 24)$	$-16(n^4 - 4n^3 + 11n^2 - 20n + 12)$	$48(n^3 - 6n^2 + 11n - 6)$
(6)(2)	$4(3n^5 - 11n^4 + 11n^3 + 119n^2 - 602n + 840)$	$-2(n^5 - 2n^4 - 27n^3 + 236n^2 - 700n + 672)$	$-(n^5 - 5n^4 - 96n^3 + 532n^2 - 672)$	$8(n^4 - 12n^3 + 62n^2 - 147n + 126)$
(5)(3)	$8(n^5 + n^4 - 39n^3 + 119n^2 + 182n - 840)$	$-4(5n^4 - 26n^3 - 25n^2 + 406n - 672)$	$-2(n^5 - 11n^4 + 61n^3 - 101n^2 - 238n + 672)$	$48(n^3 - 10n^2 + 35n - 42)$
(4) <sup>2</sup>	$n^5 - 13n^4 + 76n^3 - 37n^2 - 497n^2 - 490n + 4200$	$(-n^5 - 8n^4 + 33n^3 + 62n^2 - 868n + 1680)$	$-2(4n^4 - 31n^3 + 65n^2 + 112n - 420)$	$3(n^4 - 14n^3 + 95n^2 - 322n + 420)$
(6)(1) <sup>2</sup>	$-8(5n^4 - 2n^3 + 121n^2 - 364n + 420)$	$2(n^4 + 26n^3 + 29n^2 - 464n + 588)$	$n^4 + 56n^3 - 229n^2 + 520n - 588$	$-8(n^3 + 9n^2 - 64n + 84)$
(5)(2)(1)	$-48(2n^4 - 5n^3 - 14n^2 + 119n - 210)$	$12(n^4 + 3n^3 - 53n^2 + 211n - 294)$	$6(2n^4 - 11n^3 + 56n^2 - 197n + 294)$	$-48(n^3 - 9n^2 + 32n - 42)$
(4)(3)(1)	$-8(n^5 - 8n^4 + 81n^3 - 232n^2 + 98n + 420)$	$4(2n^4 + 9n^3 - 64n^2 - n + 294)$	$2(5n^4 - 23n^3 + 57n^2 + 15n - 294)$	$-24(n^3 - 4n^2 - 5n + 28)$
(4)(2) <sup>2</sup>	$-6(n^5 - 8n^4 + 33n^3 + 62n^2 - 868n + 1680)$	$n^5 - n^4 - 94n^3 + 868n^2 - 2952n + 3528$	$3(2n^4 - 7n^3 - 57n^2 + 380n - 588)$	$-6(n^4 - 16n^3 + 104n^2 - 305n + 336)$
(3) <sup>2</sup> (2)	$-16(4n^4 - 31n^3 + 65n^2 + 112n + 420)$	$4(2n^4 - 7n^3 - 57n^2 + 380n - 588)$	$n^5 - 12n^4 + 63n^3 - 52n^2 - 536n + 1176$	$-8(3n^3 - 33n^2 + 128n - 168)$
(5)(1) <sup>2</sup>	$48(3n^5 - 4n^4 + 31n - 42)$	$-4(3n^3 + 32n^2 - 77n - 42)$	$-2(5n^2 + 34n^2 - 117n + 126)$	$48(n^2 - 3n)$
(4)(2)(1) <sup>2</sup>	$24(n^4 + 2n^3 + 32n^2 - 155n + 210)$	$-2(n^4 + 20n^3 - 22n^2 - 119n + 210)$	$-3(19n^3 - 92n^2 + 223n - 210)$	$12(n^3 - 3n^2 + 2n)$
(3) <sup>2</sup> (1) <sup>2</sup>	$16(n^4 - 4n^3 + 50n^2 - 167n + 210)$	$-4(6n^3 + 11n^2 - 185n + 378)$	$-(n^4 + 8n^3 - 29n^2 + 176n - 476)$	$8(9n^2 - 57n + 98)$
(3)(2) <sup>2</sup> (1)	$24(n^4 + 8n^3 - 91n^2 + 322n - 420)$	$-4(n^4 + 11n^3 - 136n^2 + 526n - 672)$	$-6(n^4 - 8n^3 + 49n^2 - 166n + 224)$	$24(n^3 - 10n^2 + 38n - 49)$

$(2)^4$	$9(n^4 - 14n^3 + 95n^2 - 322n + 420)$	$-3(n^4 - 16n^3 + 104n^2 - 305n + 336)$	$-3(3n^3 - 33n^2 + 123n - 108)$	$n^4 - 18n^3 + 125n^2 - 384n + 441$
$(4)(1)^4$	$-12(n^3 + 17n^2 + 12n)$	$n^3 + 35n^2 + 138n - 504$	$41n^2 - 53n + 42$	$-6(n^2 + 7n - 28)$
$(3)(2)(1)^2$	$-96(n^3 + 2n^2 - 3n)$	$8(n^3 + 20n^2 - 87n + 126)$	$2(5n^3 + 14n^2 - 17n - 42)$	$-16(3n^2 - 14n + 21)$
$(2)^2(1)^2$	$-72(n^3 - 3n^2 + 2n)$	$18(n^3 - 7n^2 + 24n - 28)$	$3(3n^3 - 12n^2 + 15n + 14)$	$-4(n^3 - 9n^2 + 35n - 42)$
$(3)(1)^5$	$48(n^3 + 5n)$	$-4(n^3 + 23n - 42)$	$-4(n^3 + 11n - 14)$	$8(3n - 7)$
$(2)^2(1)^4$	$180(n^2 - n)$	$-3(11n^2 - 23n + 42)$	$-3(7n^2 - 11n + 14)$	$6(n^2 - 3n + 7)$
$(2)(1)^6$	$-144n$	$24n$	$16n$	$-4n$
$(1)^8$	$36$	$-6$	$-4$	$1$



A condition satisfied by the coefficients of any seminvariant is that their sum is equal to zero (See section 2). This provides another check on the entries of table II, although the seminvariant must be written in homogeneous form before the check is applied. Thus we may write

$$L_4 = \frac{n^3}{n^{(4)}} \left[ (n+1) \frac{(4)}{n} - 4(n+1) \frac{(3)(1)}{n^2} - 3(n-1) \frac{(2)^2}{n^2} + 12n \frac{(2)(1)^2}{n^3} - 6n \frac{(1)^4}{n^4} \right],$$

and the sum of coefficients is

$$(n+1) - 4(n+1) - 3(n-1) + 12n - 6n = 0.$$

Several checks arise from the fact (see section 6) that every seminvariant must be annihilated by the operator

$$(24) \quad \Omega' = \sum_{i=1}^n i s_{i-1} \frac{\partial}{\partial s_i}.$$

Another check results from the discussion of the next section and is so apparent as to need no comment.

All the checks mentioned in this section are applicable to the estimate of any seminvariant.

**7. Estimates as Sums of Simple Seminvariants.** A seminvariant such as  $L_4$  in which the coefficients of the  $m''$ 's are functions of  $n$  will be called a composite seminvariant, while a seminvariant in which the coefficients of the  $m''$ 's are purely numerical will be called simple. The fact that is to be established in this section is that every composite seminvariant is the sum of simple seminvariants. As an illustration consider  $L_4$ . It is apparent that

$$L_4 = \frac{n^4}{n^{(4)}} l_4 + \frac{n^3}{n^{(4)}} k_4,$$

where  $l_4$  and  $k_4$  are seminvariants of the sample corresponding to  $\lambda_4$  and  $\kappa_4$ . Both  $l_4$  and  $k_4$  are simple seminvariants.

That a composite seminvariant may always be expressed as a sum of simple seminvariants can be demonstrated by considering the effect of  $\Omega'$ , (24), on a composite seminvariant. The coefficients are polynomials in  $n$  and are unaffected by the operator. The expression resulting from application of the operator can vanish only if the coefficient of  $n^r$  vanishes for every  $r$ . Thus a composite seminvariant which has  $r$  different powers of  $n$  appearing in its coefficients is expressible as the sum of  $r$  simple seminvariants, which are not necessarily distinct. Table III exhibits the estimates of Thiele seminvariants of weight  $\leq 6$  as sums of simple seminvariants.

Since the factors, appearing in front of each of the simple seminvariants in the expression resulting from breaking down a composite seminvariant, are of

TABLE III  
Estimates as Sums of Seminvariants

$L_2$	$\frac{n^2}{n^{(2)}}$
$\frac{(2)}{n}$	1
$\frac{(1)^2}{n^2}$	-1

$\frac{n^3}{n^{(3)}}$
$\frac{(3)}{n}$
$\frac{(2)(1)}{n^2}$
$\frac{(1)^3}{n^3}$

$L^4$	$\frac{n^4}{n^{(4)}}$	$\frac{n^3}{n^{(3)}}$
$\frac{(4)}{n}$	1	1
$\frac{(3)(1)}{n^2}$	-4	-4
$\frac{(2)^2}{n^2}$	-3	3
$\frac{(2)(1)^2}{n^3}$	12	
$\frac{(1)^4}{n^4}$	-6	

$L_{22}$	$\frac{n^4}{n^{(4)}}$	$\frac{n^3}{n^{(3)}}$	$\frac{n^2}{n^{(2)}}$
$\frac{(4)}{n}$		1	1
$\frac{(3)(1)}{n^2}$		-4	-4
$\frac{(2)^2}{n^2}$	1	3	3
$\frac{(2)(1)^2}{n^3}$	-2		
$\frac{(1)^4}{n^4}$	1		

$L_{22}$	$\frac{n^3}{n^{(3)}}$	$\frac{n^4}{n^{(4)}}$	$\frac{n^2}{n^{(2)}}$
$\frac{(5)}{n}$		1	1
$\frac{(4)(1)}{n^2}$		-5	-5
$\frac{(3)(2)}{n^2}$	1	2	2
$\frac{(3)(1)^2}{n^3}$	-1	8	8
$\frac{(2)^2(1)}{n^3}$	-3	-6	-6
$\frac{(2)(1)^3}{n^4}$	5		
$\frac{(1)^5}{n^5}$	-2		

$L_6$	$\frac{n^6}{n^{(6)}}$	$\frac{2n^5}{n^{(5)}}$	$\frac{n^4}{n^{(4)}}$	$\frac{4n^3}{n^{(3)}}$
$\frac{(6)}{n}$	1	8	11	1
$\frac{(5)(1)}{n^2}$	-6	-48	-66	-6
$\frac{(4)(2)}{n^2}$	-15	-15	105	15
$\frac{(3)^2}{n^2}$	-10	10	-50	-10
$\frac{(4)(1)^2}{n^3}$	30	135	60	
$\frac{(3)(2)(1)}{n^3}$	120		-120	
$\frac{(2)^3}{n^3}$	30	-45	60	
$\frac{(3)(1)^3}{n^4}$	-120	-180		
$\frac{(2)^2(1)^2}{n^4}$	-270	135		
$\frac{(2)(1)^4}{n^5}$	360			
$\frac{(1)^6}{n^6}$	-120			

$L_6$	$\frac{n^5}{n^{(5)}}$	$\frac{5n^4}{n^{(4)}}$
$\frac{(5)}{n}$	1	1
$\frac{(4)(1)}{n^2}$	-5	-5
$\frac{(3)(2)}{n^2}$	-10	2
$\frac{(3)(1)^2}{n^3}$	20	8
$\frac{(2)^2(1)}{n^3}$	30	-6
$\frac{(2)(1)^3}{n^4}$	-60	
$\frac{(1)^5}{n^5}$	24	

TABLE III—Continued

$L_{12}$	$\frac{n^6}{n^{(6)}}$	$\frac{n^5}{n^{(5)}}$	$\frac{2n^4}{n^{(4)}}$	$\frac{n^3}{n^{(3)}}$	$\frac{4n^2}{n^{(2)}}$
$(6) \frac{1}{n}$		1	1	7	1
$(5)(1) \frac{1}{n^2}$		-6	-6	-42	-6
$(4)(2) \frac{1}{n^2}$	1	15	5	45	15
$(3)^2 \frac{1}{n^2}$		-4	5	-10	-10
$(4)(1)^2 \frac{1}{n^3}$	-1		10	60	
$(3)(2)(1) \frac{1}{n^3}$	-4	24	-50	-120	
$(2)^2 \frac{1}{n^3}$	-3	-21	30	60	
$(3)(1)^2 \frac{1}{n^4}$	4	-36	20		
$(2)^2(1)^2 \frac{1}{n^4}$	15	27	-15		
$(2)(1)^4 \frac{1}{n^5}$	-18				
$(1)^6 \frac{1}{n^6}$	6				

$L_{22}$	$\frac{n^6}{n^{(6)}}$	$\frac{n^5}{n^{(5)}}$	$\frac{n^4}{n^{(4)}}$	$\frac{n^3}{n^{(3)}}$	$\frac{5n^2}{n^{(2)}}$	$\frac{4n^2}{n^{(2)}}$
$(6) \frac{1}{n}$		1	2	1	1	1
$(5)(1) \frac{1}{n^2}$		-6	-12	-6	-6	-6
$(4)(2) \frac{1}{n^2}$		-6	-15	3	15	15
$(3)^2 \frac{1}{n^2}$	1	8	25	2	-10	-10
$(4)(1)^2 \frac{1}{n^3}$		21	45	12		
$(3)(2)(1) \frac{1}{n^3}$	-6	-24	-90	-24		
$(2)^2 \frac{1}{n^3}$		9	45	12		
$(3)(1)^2 \frac{1}{n^4}$	4	-12				
$(2)^2(1)^2 \frac{1}{n^4}$	9	9				
$(2)(1)^4 \frac{1}{n^5}$	-12					
$(1)^6 \frac{1}{n^6}$	4					

$L_{32}$	$\frac{n^6}{n^{(6)}}$	$\frac{n^5}{n^{(5)}}$	$\frac{n^4}{n^{(4)}}$	$\frac{n^3}{n^{(3)}}$	$\frac{5n^2}{n^{(2)}}$	$\frac{4n^2}{n^{(2)}}$
$(6) \frac{1}{n}$						
$(5)(1) \frac{1}{n^2}$						
$(4)(2) \frac{1}{n^2}$						
$(3)^2 \frac{1}{n^2}$						
$(4)(1)^2 \frac{1}{n^3}$						
$(3)(2)(1) \frac{1}{n^3}$						
$(2)^2 \frac{1}{n^3}$						
$(3)(1)^2 \frac{1}{n^4}$						
$(2)^2(1)^2 \frac{1}{n^4}$						
$(2)(1)^4 \frac{1}{n^5}$						
$(1)^6 \frac{1}{n^6}$						

successively lower order with respect to  $n$ ; it is possible to obtain approximations of various orders to the value of an estimate by using the appropriate portion of the expression given in the table.

**8. The Estimates of the  $\kappa$ 's.** The seminvariant  $\kappa_r$  possesses an interesting property which will be called invariance under estimate. By this is meant that the estimate of  $\kappa_r$  is  $k_r$  multiplied by a suitable factor. In particular,  $\kappa_2 = \mu_2$  and  $\kappa_3 = \mu_3$  and it is well known that

$$E^{-1}[\mu_2] = \frac{n^2}{n^{(2)}} m_2, \quad E^{-1}[\mu_3] = \frac{n^3}{n^{(3)}} m_3$$

so that the  $\kappa_r$  certainly possesses the property for  $r = 2$  and 3. It can be shown, however, that

$$(25) \quad K_{2r} = \frac{n^2}{n^{(2)}} K_{2r}, \quad K_{2r+1} = \frac{n^3}{n^{(3)}} K_{2r+1}.$$

From (15)

$$\kappa_{2r} = \frac{1}{2} \sum_{i=0}^{2r} \binom{2r}{i} \mu_i' \mu_{2r-i}'$$

so that

$$K_{2r} = \frac{1}{2} \sum_{i=1}^{2r-1} (-1)^i \frac{(i, 2r-i)}{n^{(2)}} + \frac{(2r)}{n}.$$

By the Binet-Waring identities [15; 6-7]

$$(26) \quad (a \cdot b) = (a)(b) - (a+b)$$

and this holds for power products regardless of the values of  $a$  and  $b$ . Hence

$$\begin{aligned} K_{2r} &= \frac{(2r)}{n} + \frac{1}{2} \sum_{i=1}^{2r-1} (-1)^i \binom{2r}{i} \frac{(i)(2r-i)}{n^{(2)}} \\ &= \frac{(2r)}{n} \left[ 1 - \frac{1}{2} \sum_{i=1}^{2r-1} \frac{(-1)^i \binom{2r}{i}}{n-1} \right] + \frac{1}{2} \sum_{i=1}^{2r-1} (-1)^i \binom{2r}{i} \frac{(i)(2r-i)}{n^{(2)}}. \end{aligned}$$

Since

$$\sum_{i=0}^{2r} (-1)^i \binom{2r}{i} = 0 = 2 + \sum_{i=1}^{2r-1} (-1)^i \binom{2r}{i},$$

the coefficient of  $\frac{(2r)}{n}$  above is  $\frac{n}{n-1}$  and it follows immediately that

$$K_{2r} = \frac{1}{2} \sum_{i=0}^{2r} (-1)^i \binom{2r}{i} \frac{(i)(2r-i)}{n^{(2)}} = \frac{n^2}{n^{(2)}} K_{2r}.$$

This proves the first half of (25) and the second half can be proved in similar fashion, although with considerably more difficulty.

9. **Other Simple Seminvariants which are Invariant under Estimate.** It has been previously remarked (Chapter I, section 2) that the  $\kappa$  system of seminvariants are the seminvariants of minimum degree, those of even weight being of second degree and those of odd weight being of third degree. The  $\kappa_r$ 's are the only seminvariants of degree 2, but for odd weights greater than 7, there exist more than one seminvariant of degree 3. It is not difficult to show that these additional minimum degree seminvariants are also invariant under estimate. The type of proof used could have been applied equally well to obtain the results of the preceding section and indicates that the property of invariance under estimate which is possessed by the  $\kappa$ 's is a direct result of their minimum degree property.

Consider the estimate in power product form of any seminvariant of degree 3 and odd weight. Power products of 1, 2 and 3 parts will appear. By the Binet-Waring identities each three part power product  $(abc)$  yields a third degree power sum product  $(a)(b)(c)$  plus other products of lower degree. Since  $(a)(b)(c)$  comes only from  $(abc)$  its coefficient must be identical with that of  $(abc)$  and will therefore be a constant divided by  $n^{(3)}$ . The coefficient of each second degree product of power sums will be a sum of terms, the first of which comes from the corresponding two part power product with a coefficient identical with that of the power product, and the others come from the three part power products. Then the coefficient of a second degree product of power sums must be of the form

$$\frac{c_1}{n^{(2)}} + \frac{c_2 + c_3 + \dots + c_t}{n^{(2)}} = \frac{c_1 n + c_2'}{n^{(2)}}.$$

Similarly the coefficient of the first degree power sum term will be of the form

$$\frac{d_1 n^2 + d_2 n + d_3}{n^{(3)}}.$$

Since the estimate of a seminvariant is a seminvariant, it follows that  $d_3 = 0$ . This is true because the coefficient of  $\frac{(r-1)(1)}{n^2}$  must be the coefficient of  $\frac{(r)}{n}$  multiplied by  $-r$ . Furthermore  $c_2' = d_2 = 0$  for if the contrary be assumed it is immediately possible to break the composite seminvariant into two simple seminvariants, the first being of degree 3 (the original seminvariant) and the second of degree 2. Since for odd weights no seminvariant of degree 2 exists, it follows that any seminvariant of degree 3 and odd weight is invariant under estimate. It is also apparent that the factor  $n^3/n^{(3)}$  must appear in the estimate.

10. **Composite Seminvariants which are Invariant under Estimate.** For each weight  $r \geq 4$  there exists a composite seminvariant which is invariant under estimate. For weights 4 and 5 this seminvariant is easily obtained by use

of Table III. Thus for weight 4, form the seminvariant  $\lambda_4 + c_{22}\lambda_2^2$ . From the table we find that

$$\begin{aligned} E^{-1}[\lambda_4 + c_{22}\lambda_2^2] &= \frac{n^4}{n^{(4)}} l_4 + \frac{n^3}{n^{(4)}} k_4 + c_{22} \frac{n^4}{n^{(4)}} l_2^2 - c_{22} \frac{n}{(n-2)^{(2)}} k_4 \\ &= \frac{n^4}{n^{(4)}} (l_4 + c_{22} l_2^2) + \frac{n}{n^{(4)}} (n^2 - n^{(2)} c_{22}) k_4. \end{aligned}$$

If  $c_{22} = n^2/n^{(2)}$  the seminvariant is invariant under estimate. This seminvariant is

$$(27) \quad \psi_4 = \lambda_4 + \frac{n^2}{n^{(2)}} \lambda_2^2.$$

In similar fashion we find for weight 5

$$(28) \quad \psi_5 = \lambda_5 + \frac{5n^2}{n^{(2)}} \lambda_3 \lambda_2.$$

For weights  $> 5$  considerably more difficulty is encountered. For weight 6, for example, we consider the seminvariant

$$\lambda_6 + c_{42} \lambda_4 \lambda_2 + c_{33} \lambda_3^2 + c_{222} \lambda_2^3.$$

By use of table III we obtain

$$E^{-1}[\lambda_6 + c_{42} \lambda_4 \lambda_2 + c_{33} \lambda_3^2 + c_{222} \lambda_2^3] = \frac{n^6}{n^{(6)}} (l_6 + c_{42} l_4 l_2 + c_{33} l_3^2 + c_{222} l_2^3) + \Phi,$$

where  $\Phi$  is a sum of other seminvariants with coefficients which are functions of  $n$  and  $c_{42}$ ,  $c_{33}$ ,  $c_{222}$ . Now there are only four linearly independent seminvariants of weight 6 and it is necessary that one of these involve the term  $(1)^6/n^6$ . By an argument analogous to that of the previous section this term cannot appear in  $\Phi$  and therefore  $\Phi$  is expressible in terms of three or fewer seminvariants. Actually three are necessary and equating the coefficients of these to zero the values of  $c_{42}$ ,  $c_{33}$  and  $c_{222}$  are uniquely determined. The result is somewhat lengthy and scarcely of sufficient interest to record here.

The same sort of procedure can be used for determining seminvariants of higher order which are invariant under estimate, but the labor of computation becomes very great.

It is possible to obtain moment functions which are invariant under estimate by means of a set of equations given by Dwyer [13; 38-39]. These equations connect the coefficients of a general isobaric moment function and the coefficients of the expected value of that function. In his notation if, for example,

$$f_4 = a_4(4) + 4a_{31}(3)(1) + 3a_{22}(2)^2 + 6a_{211}(2)(1)^2 + a_{1111}(1)^4,$$

then

$$E[f_4] = b_4 n \mu'_4 + 4b_{31} n^{(2)} \mu'_3 \mu'_1 + 3b_{22} n^{(2)} \mu'^2_2 + 6b_{211} n^{(3)} \mu'_2 \mu'^2_1 + n^{(4)} b_{1111} \mu'^4_1,$$

wherein:

$$\begin{aligned}
 (29) \quad & a_4 + 4a_{31} + 3a_{22} + 6a_{211} + a_{1111} = b_4, \\
 & a_{31} + 3a_{211} + a_{1111} = b_{31}, \\
 & a_{22} + 2a_{211} + a_{1111} = b_{22}, \\
 & a_{211} + a_{1111} = b_{211}, \\
 & a_{1111} = b_{1111}.
 \end{aligned}$$

The problem at hand demands that

$$\begin{aligned}
 E \left[ na_4 \frac{(4)}{n} + 4n^2 a_{31} \frac{(3)(1)}{n^2} + 3n^2 a_{22} \frac{(2)^2}{n^2} + 6n^3 a_{211} \frac{(2)(1)^2}{n^3} + n^4 a_{1111} \frac{(1)^4}{n^4} \right] \\
 = \lambda [na_4 \mu'_4 + 4n^{(2)} a_{31} \mu'_3 \mu'_1 + 3n^{(2)} a_{22} \mu'^2_2 + 6n^{(3)} a_{211} \mu'_2 \mu'^2_1 + n^{(4)} a_{1111} \mu'^4_1]
 \end{aligned}$$

so that the equations (29) become

$$\begin{aligned}
 n^4 a_{1111} &= \lambda n^{(4)} a_{1111}, \\
 n^3 a_{211} &= \lambda n^{(3)} (a_{211} + a_{1111}), \\
 n^2 a_{22} &= \lambda n^{(2)} (a_{22} + 2a_{211} + a_{1111}), \\
 n^2 a_{31} &= \lambda n^{(2)} (a_{31} + 3a_{211} + a_{1111}), \\
 na_4 &= \lambda n (a_4 + 4a_{31} + 3a_{22} + 6a_{211} + a_{1111}),
 \end{aligned}$$

and from these equations  $a_4$ ,  $a_{31}$ ,  $a_{22}$ ,  $a_{211}$  can be found in terms of  $a_{1111}$ . Obviously there is only one solution if none of the  $a$ 's are zero. In general, for any weight  $r$ , a similar system of equations can be found and they determine the coefficients of a moment function of weight  $r$  which is invariant under estimate. It appears that this moment function is always a seminvariant although no proof of the fact has been found. The moment functions of weight 4, 5 and 6 obtained by this method are identical with  $\psi_4$ ,  $\psi_5$  and  $\psi_6$  defined above.

**Conclusion.** The results of this paper include:

1. A demonstration of the fact that the theory of statistical seminvariants is identical with the theory of algebraic seminvariants.
2. The introduction of new statistical seminvariants.
3. Simplification of the computation of estimates.
4. Proof that the estimate of any seminvariant is also a seminvariant.
5. Proof of the existence of a trio of seminvariants with the same numerical coefficients.
6. A discussion of seminvariants which are invariant under estimate.

Many thanks are due Professor P. S. Dwyer for his able guidance in the preparation of this paper and to Professors C. C. Craig and J. A. Nyswander for helpful comments.

## REFERENCES

- [1] L. E. DICKSON, *Algebraic Invariants* (1914).
- [2] E. B. ELLIOTT, *An Introduction to the Algebra of Quantics* (1895). Reference is to the second edition (1913).
- [3] J. HAMMOND, "On the Solution of the Differential Equation of Sources," *Amer. Jour. of Math.*, Vol. 5 (1882), pp. 218-227.
- [4] P. A. MACMAHON, "Seminvariants and Symmetric Functions," *Amer. Jour. of Math.*, Vol. 6 (1884), pp. 131-163.
- [5] BURNSIDE and PANTON, *Theory of Equations* (1881). Reference is to the 1904 edition, Vol. II.
- [6] P. S. DWYER, "Combined Expansions of Products of Symmetric Power Sums and Sums of Symmetric Power Products with Applications to Sampling." Part I. *Annals of Math. Stat.*, Vol. IX, 1, (1938), pp. 1-47. Part II, Vol. IX, 2, (1938), pp. 97-132.
- [7] T. N. THIELE, *Theory of Observations* (1903).
- [8] N. P. BERTILSEN, "On the Compatibility of Frequency Constants and on Presumptive Laws of Errors." *Skandinavisk Aktuarietidskrift*, Vol. 10 (1927), pp. 129-156.
- [9] A. CAYLEY, "A Memoir on Seminvariants." *Amer. Jour. of Math.* Vol. 7 (1885), pp. 1-25.
- [10] T. N. THIELE, *Almindelig Iagttagelseslaere* (1889).
- [11] A. A. TSCHUPROW, *Grundbegriffe und Grundprobleme der Korrelationstheorie* (1925).
- [12] R. A. FISHER, "Moments and Product Moments of Sampling Distributions." *Proc. London Math. Soc.*, Vol. 2 (30), (1929), pp. 199-238.
- [13] P. S. DWYER, "Moments of Any Rational Integral Isobaric Sample Moment Function." *Annals of Math. Stat.*, Vol. 8 (1937), pp. 21-65.
- [14] J. R. ROE, "Interfunctional Expressibility Tables of Symmetric Functions." Distributed by Syracuse University (1931).
- [15] FAÁ DE BRUNO, *Theorie des Formes Binaires* (1876).

UNIVERSITY OF MICHIGAN,  
ANN ARBOR, MICHIGAN.



# THE ERRORS INVOLVED IN EVALUATING CORRELATION DETERMINANTS

BY PAUL G. HOEL

1. **Introduction.** Many statistical problems require for their solution the evaluation of correlation determinants. The method usually employed for such evaluation is that of Chio,<sup>1</sup> in which the order of the determinant is reduced by successive operations with selected pivotal elements. The repeated multiplications and subtractions involved in the method necessitate rounding off the elements in the successively reduced determinants. The calculated value of the original determinant is therefore in error; and so the question naturally arises as to the magnitude of this error.

Previous attempts to answer this question seem to be satisfied with finding an upper bound for the magnitude of the difference between the value of the original determinant and its value after its elements have been rounded off. Moreover, this bound is expressed in terms of the errors in the elements and the minors of the original determinant, whose values are assumed to be known exactly from calculation. However, several reductions are often needed before the value of the determinant can be obtained; and furthermore the minors are subject to the same type of errors as the determinant itself. The problem, therefore, is to find an upper bound for the magnitude of the difference between the final calculated value of the determinant and the determinant itself which involves only calculated quantities.

This paper treats the problem from two different points of view. In the first part an upper bound is obtained for the magnitude of the error. In the second part the first order error terms are given more detailed consideration, with the result that an upper probability bound is obtained for the error.

2. **Absolute Bounds.** Consider the correlation determinant  $\Delta = |r_{ij}|$ . To evaluate  $\Delta$  by the method of Chio, it is convenient to select diagonal elements as pivots. It will be assumed without loss of generality that the upper left diagonal element is always chosen as the pivotal element in each reduction. After each reduction, elements are rounded off to a fixed decimal accuracy. Let  $a_{ij}^k$  represent the element  $i, j$  after the  $k$ -th reduction,  $x_{ij}^k$  the difference between the rounded value of element  $a_{ij}^k$  and  $a_{ij}^k$  itself. After  $k$  reductions, we arrive at the determinant

$$F^k = \begin{vmatrix} a_{k+1, k+1}^k + x_{k+1, k+1}^k & & \\ & \ddots & \\ & & a_{nn}^k + x_{nn}^k \end{vmatrix}$$

<sup>1</sup> See for example, Whittaker and Robinson *Calculus of Observations*, p. 71.

By treating  $F^k$  as a function of the  $x^k$ , it may be expanded by Taylor's formula as follows:

$$(1) \quad F^k = A^k + \sum_{i,j=k+1}^n x_{ij}^k A_{ij}^k + \frac{1}{2!} \sum_{k+1}^n \sum_{k+1}^n x_{ij}^k x_{pq}^k A_{ijpq}^k +$$

where  $A^k$  is the value of  $F^k$  for all  $x^k$  zero,  $A_{ij}^k$  is the cofactor of  $a_{ij}^k$  in  $A^k$ , etc.

For a determinant of order  $n$ , the value of the determinant obtained after a single reduction is the value of the original determinant multiplied by the  $n - 2$  power of the pivotal element used. Applying this to  $F^k$ , it follows that

$$\begin{aligned} A^k &= (a_{kk}^{k-1} + x_{kk}^{k-1})^{n-k-1} F^{k-1} = H_k^{n-k-1} F^{k-1} \\ A_{ij}^k &= H_k^{n-k-2} F_{ij}^{k-1} \\ A_{ijpq}^k &= H_k^{n-k-3} F_{ijpq}^{k-1}, \end{aligned}$$

etc., where the exponents of  $H_k$  are ordinary exponents rather than notation. Substituting in (1),

$$F^k = H_k^{n-k-1} F^{k-1} + H_k^{n-k-2} \sum_{k+1}^n x_{ij}^k F_{ij}^{k-1} + \frac{1}{2!} H_k^{n-k-3} \sum_{k+1}^n \sum_{k+1}^n x_{ij}^k x_{pq}^k F_{ijpq}^{k-1} + \dots$$

In order to express  $F^k$  in terms of the original determinant, this expansion will be condensed by means of the following operational notation.

$$(2) \quad F^k = (1 + D + D^2 + \dots + D^{n-k}) H_k^{n-k-1} F^{k-1},$$

where  $D^i$  operates on  $H_k^{n-k-1} F^{k-1}$  by reducing the exponent of  $H_k^{n-k-1}$  by  $i$  units, by summing from  $k + 1$  to  $n$  the product of  $i$  terms in  $x^k$  with the corresponding cofactors of  $F^{k-1}$ , and dividing the result by factorial  $i$ . Using this as a recursion formula,

$$\begin{aligned} F^k &= (1 + D + \dots + D^{n-k}) H_k^{n-k-1} (1 + \dots + D^{n-k+1}) H_{k-1}^{n-k} \dots \\ &\quad (1 + \dots + D^{n-1}) H_1^{n-2} F^0. \end{aligned}$$

However,

$$F^0 = \begin{vmatrix} a_{11} + x_{11} & & \\ & \ddots & \\ & & a_{nn} + x_{nn} \end{vmatrix} = \Delta,$$

since we assume that  $x_{ij} = 0$  for our original determinant. Consequently,

$$(3) \quad \begin{aligned} F^k &= (1 + \dots + D^{n-k}) H_k^{n-k-1} (1 + \dots + D^{n-k+1}) H_{k-1}^{n-k} \dots \\ &\quad (1 + \dots + D^{n-1}) H_1^{n-2} \Delta. \end{aligned}$$

Since  $D^i$  operates on  $F^{k-1}$  in (2) to extract the proper cofactor of  $i$  less rows than in  $F^{k-1}$ , which in turn reduces the exponent of all factors  $H_{k-1}$  in the expansion of  $F^{k-1}$  by  $i$  units,  $D^i$  reduces the exponent of all  $H$ 's following it in the expansion of  $F^k$  in (3) by  $i$  units.

Following these rules of operation, and expanding so as to collect terms of the same degree in the  $x$ 's, we may write

$$(4) \quad F^k = H_k^{n-k-1} \dots H_1^{n-2} \Delta + H_k^{n-k-2} \dots H_1^{n-3} (\text{terms in } x_{ij}) + \\ H_k^{n-k-3} \dots H_1^{n-4} (\text{terms in } x_{ij}x_{pq}) + \dots$$

Letting  $H = H_k H_{k-1} \dots H_1$  and  $C = H_k^{n-k-1} \dots H_1^{n-2}$ , we may write

$$I = F^k - C\Delta = C \left[ \frac{1}{H} (\text{terms in } x_{ij}) + \frac{1}{H^2} (\text{terms in } x_{ij}x_{pq}) + \dots \right];$$

and hence

$$(5) \quad J = \frac{F^k}{C} - \Delta = \frac{1}{H} (\text{terms in } x_{ij}) + \frac{1}{H^2} (\text{terms in } x_{ij}x_{pq}) + \dots$$

Now  $J$  is the difference between the calculated value of  $\Delta$ , using Chio's reduction method and rounding off after each reduction, and the true value of  $\Delta$ . We are interested in finding an upper bound for the magnitude of  $J$ . To accomplish this we shall first overestimate the number of terms in the various sums of (5), then find an upper bound for the magnitude of the terms in these sums, and finally combine the two results.

In counting terms by means of (3), we may ignore the  $H$ 's since they merely serve as coefficients of the  $x$ 's. Therefore consider the nature of the terms in

$$(1 + \dots + D^{n-k})(1 + \dots + D^{n-k+1}) \dots (1 + \dots + D^{n-1})\Delta.$$

Now  $(1 + \dots + D^r)\Delta$  contains the sums  $\sum_{n-s+1}^n x_{ij}\Delta_{ij}, \frac{1}{2!} \sum_{n-s+1}^n \sum_{n-s+1}^n x_{ij}x_{pq}\Delta_{ijpq}$ , etc.;

hence it contains  $s^2$  terms in  $x_{ij}$ ,  $\frac{s^2(s-1)^2}{2}$  terms in  $x_{ij}x_{pq}$ , etc. Each of these is not greater than  $s^2, {}_sC_2$ , etc.; consequently, the number of terms of each type is not greater than the coefficient of the corresponding power of  $D$  in the expansion of  $(1 + D)^{s^2}$ . Therefore,

$$(6) \quad (1 + D)^{(n-k)^2} (1 + D)^{(n-k+1)^2} \dots (1 + D)^{(n-1)^2} = (1 + D)^m,$$

where  $m = (n-k)^2 + \dots + (n-1)^2$ , contains at least as many terms of each type as are found in the expansion of  $F^k$ . This gives us the desired overestimate of the number of terms in the various sums of (5).

In finding upper bounds for the magnitudes of terms, it is to be noted that (4) is written with all common factors extracted from each set of terms of the same degree in the  $x$ 's. In the parenthesis containing terms consisting of the product of  $r$   $x$ 's, the first sum will have unity for its coefficient while the last sum will have  $H_k^r H_{k-1}^r \dots H_2^r$  as coefficient, with all sums between having as coefficients products of  $H$ 's with exponents  $\leq r$ . Hence an upper bound for all coefficients in this parenthesis may be written as  $\bar{H}^r$ , where  $\bar{H}$  is the magnitude of the product of those  $H$ 's whose magnitude is greater than unity, but unity if none exceeds

unity. Now terms in  $x_{ij}$  are multiplied by  $\Delta_{ij}$ , those in  $x_{ij}x_{pq}$  by  $\Delta_{ijpq}$ , etc.; therefore let  $\bar{\Delta}_{ij}$ ,  $\bar{\Delta}_{ijpq}$ , etc., be the absolute values of the largest in magnitude of such cofactors. With this notation for upper bounds for magnitudes of terms, and (6) giving an upper bound for the number of terms, we may write an upper bound for the magnitude of  $J$  as follows:

$$(7) \quad |J| \leq \left(\frac{H}{H}\epsilon\right) {}_m C_1 \bar{\Delta}_{ij} + \left(\frac{H}{H}\epsilon\right)^2 {}_m C_2 \bar{\Delta}_{ijpq} + \dots,$$

where  $\epsilon \geq |x|$  is the maximum error of rounding. This result is valid for any determinant with real elements. All quantities on the right are available from calculations except the  $\bar{\Delta}$ ; consequently this upper bound will be useful only if satisfactory bounds exist for the minors of the determinant. It can be shown that (7) holds for any minor of  $\Delta$ , say  $\Delta_{uv}$ , if the  $\bar{\Delta}$  have  $uv$  added as subscripts; and therefore it may be applied to the question of the accuracy of least square solutions.

For the correlation determinant  $\Delta$  it can be shown that the magnitude of a minor of order  $n - k$  is bounded by  $k!/2^{1k}$  for  $k$  even and  $k!/2^{1(k-1)}$  for  $k$  odd.

Setting  $a = \frac{H}{H}\epsilon$  and substituting these bounds in (7),

$$\begin{aligned} |J| &\leq am + a^2 {}_m C_2 \frac{2!}{2} + a^3 {}_m C_3 \frac{3!}{2} + a^4 {}_m C_4 \frac{4!}{2^2} + \dots \\ &\leq am + \frac{a^2 m^2}{2} + \frac{a^3 m^3}{2} + \frac{a^4 m^4}{2^2} + \dots \\ (8) \quad &\leq am + \frac{a^2 m^2}{2} + \frac{a^3 m^3}{2(1-am)}, \end{aligned}$$

for  $am < 1$ . Since  $am$  is obtainable from the calculations for  $\Delta$ , this is the desired upper bound for the error in question.

**3. Probability Bounds.** In order to find probability bounds for this error, it will be necessary to expand the  $H$ 's since they involve the variables  $x$ . Consider  $H_k = a_{kk}^{k-1} + x_{kk}^{k-1}$ . Since  $a_{kk}^{k-1}$  came from repeated reductions of  $\Delta$ , it is expressible in terms of the  $x$ 's and the minors of  $\Delta$ . To obtain this expansion of  $H_k$  consider

$$G^s = \begin{vmatrix} a_{k-s+1, k-s+1}^{k-s} + x_{k-s+1, k-s+1}^{k-s} & & \\ & \ddots & \\ & & a_{kk}^{k-s} + x_{kk}^{k-s} \end{vmatrix}$$

Using the same methods as for  $F^k$ , this may be written as

$$G^s = B^s + \sum_{k-s+1}^k x_{ij}^{k-s} B_{ij}^s + \frac{1}{2!} \sum_{k-s+1}^k \sum_{k-s+1}^k x_{ij}^{k-s} x_{pq}^{k-s} B_{ijpq}^s + \dots,$$

where  $B^s$  is the value of  $G^s$  for all  $x^{k-s}$  zero, etc., and where  $B^s = H_{k-s}^{-1}G^{s+1}$ ,  $B_{ij}^s = H_{k-s}^{-2}G_{ij}^{s+1}$ , etc. Substituting,

$$G^s = H_{k-s}^{-1}G^{s+1} + H_{k-s}^{-2} \sum x_{ij}^{k-s} G_{ij}^{s+1} + \frac{1}{2!} H_{k-s}^{-3} \sum \sum x_{ij}^{k-s} x_{pq}^{k-s} G_{ijpq}^{s+1} + \dots$$

Using operational notation here also, this may be written as

$$G^s = (1 + E + E^2 + \dots + E^s) H_{k-s}^{-1} G^{s+1},$$

where the  $E$ 's operate the same as the  $D$ 's, except that sums are taken from  $k - s + 1$  to  $k$  rather than from  $n - s + 1$  to  $n$ . Treating this as a recursion formula,

$$H_k = G^1 = (1 + E) H_{k-1}^0 (1 + E + E^2) H_{k-2}^1 \dots (1 + \dots + E^{k-1}) H_1^{k-2} G^k.$$

However,

$$G^k = \begin{vmatrix} a_{11} + x_{11} & & \\ & \ddots & \\ & & a_{kk} + x_{kk} \end{vmatrix} = \begin{vmatrix} a_{11} & & \\ & \ddots & \\ & & a_{kk} \end{vmatrix} = \Delta_k.$$

Consequently,

$$(9) \quad H_k = (1 + E) H_{k-1}^0 (1 + E + E^2) H_{k-2}^1 \dots (1 + \dots + E^{k-1}) H_1^{k-2} \Delta_k.$$

Since the  $E$ 's operate on the following  $H$ 's to reduce their exponents, the number of terms of various types, that is, of various degrees in the  $x$ 's, will not be decreased if the order of  $H$ 's is disregarded and their exponents held fixed. Therefore consider

$$(10) \quad H'_k = (1 + E)(1 + E + E^2) \dots (1 + \dots + E^{k-1}) \Delta_k H_{k-1}^0 \dots H_1^{k-2}$$

as an ordinary recursion formula in the  $H$ 's for overestimating the number of terms of various types. If (10) is substituted for successive  $H$ 's within itself in a systematic manner until no  $H$ 's remain, it will be found that

$$(11) \quad H'_k = (1 + E) \dots (1 + \dots + E^{k-1}) \Delta_k \\ [(1 + E) \dots (1 + \dots + E^{k-3}) \Delta_{k-2}]^{2^0} \dots [(1 + E) \Delta_2]^{2^{k-4}} [\Delta_1]^{2^{k-5}}.$$

To merely count terms it is permissible to combine like terms to give

$$H'_k = (1 + E)^{2+2+2+\dots+2^{k-4}} (1 + E + E^2)^{2+2+2+\dots+2^{k-5}} \dots (1 + \dots + E^{k-1}) K \\ = (1 + E)^{2^{k-3}} (1 + E + E^2)^{2^{k-4}} \dots (1 + \dots + E^{k-1}) K,$$

where  $K$  is the product of the  $\Delta$ 's. Since the  $E$ 's operate like the  $D$ 's, the same arguments as those used to arrive at (6) may be used to replace  $(1 + E + \dots + E^s)$  by  $(1 + E)^s$  for overestimating the number of terms. Hence, the number of terms of various types in  $H_k$  is not greater than those in

$$(1 + E)^{2^{k-3}} (1 + E)^{2^{k-4}} \dots (1 + E)^{(k-2) \cdot 2^0} (1 + E)^{(k-1) \cdot 1} = (1 + E)^{w_k},$$

where  $w_k = 2^{k-3} + 2^2 \cdot 2^{k-4} + \dots + (k-2)^2 \cdot 2^0 + (k-1)^2$ . Therefore the number of terms of various types in  $H_k^{n-k-1} \dots H_1^{n-2}$  is not greater than in

$$(12) \quad (1 + E)^{(n-k-1)w_k + (n-k)w_{k-1} + \dots + (n-2)w_1} = (1 + E)^t.$$

It is easily shown that  $t$  can be condensed into the form

$$(13) \quad t = [2^{k-2}(n-k)-1] + 2^2[2^{k-3}(n-k)-1] + \dots + (k-1)^2[2^0(n-k)-1].$$

From (3) it is evident that the number of terms of various types in  $F^k$  will not be greater than those in the expansion of  $F^k$  when the exponents of the  $H$ 's are held fixed. But from (6) we have an upper bound for the number of terms arising from the  $D$ 's, and from (12) those arising from the  $H$ 's; hence the number of terms in question will certainly be bounded by those in

$$(14) \quad (1 + D)^{m+t} = (1 + D)^u.$$

Now consider the magnitude of terms. The terms arising from the operation of  $D$ 's contain minors of  $\Delta$  as factors, while those arising from the operation of  $E$ 's contain minors of  $\Delta_i$ , where  $i$  ranges from 1 to  $k$ . Let  $\Delta'_{ij}$ , etc., denote an upper bound for the magnitudes of all such minors of the same number of subscripts. It is easily shown that  $\Delta'$  with  $2r$  subscripts is not less than the magnitude of the product of several minors whose subscripts total  $2r$  in number. The terms of various types also contain as factors products of the constant terms in the  $H$ 's. The constant term in  $H_k$ , which will be denoted by  $h_k$ , can be obtained from (11) by operating with all ones since it will be unaffected by disregarding the order of operation. Hence,

$$h_k = \Delta_k \Delta_{k-2} \Delta_{k-3}^2 \dots \Delta_2^{k-4} \Delta_1^{k-3}.$$

Since the  $\Delta_i$  are principal minors of a positive definite determinant with no element greater than unity,  $h_k$  has unity for an upper bound. Thus, an upper bound for the magnitude of any term in the product of  $i$   $x$ 's will be  $\epsilon^i$  times  $\Delta'$  with  $2i$  subscripts.

With upper bounds now available for the number of terms and the magnitudes of terms, we are in a position to consider the complete expansion of  $I$  in which the coefficients of the  $x$ 's will be constants rather than  $H$ 's. Evidently the terms in  $x_{ij}$  will come from the terms in  $x_{ij}$  of (4) with the  $H$ 's replaced by the constant terms in their expansions. If  $Z$  denotes these terms, then

$$(15) \quad Z = h_k^{n-k-2} \dots h_1^{n-3} \left[ \sum_{k+1}^n x_{ij}^k \Delta_{ij} + h_k \sum_k^n x_{ij}^{k-1} \Delta_{ij} \right. \\ \left. + \dots + h_k \dots h_2 \sum_2^n x_{ij}^1 \Delta_{ij} \right].$$

Now consider an upper bound for  $|I - Z|$ . Since  $I - Z$  involves only terms in the product of two or more  $x$ 's, we need consider an upper bound for such terms only. From the results of the two preceding paragraphs, we obtain

$$|I - Z| \leq \epsilon^2 {}_\mu C_2 \Delta'_{ijpq} + \epsilon^3 {}_\mu C_3 \Delta'_{ijpquv} + \dots$$

But from the paragraph containing (8), bounds are available for the  $\Delta'$ ; hence

$$\begin{aligned} |I - Z| &\leq \epsilon^2 {}_\mu C_2 + \epsilon^3 {}_\mu C_3 \frac{3!}{2} + \epsilon^4 {}_\mu C_4 \frac{4!}{2^2} + \dots \\ &\leq \frac{\epsilon^2 \mu^2}{2} + \frac{\epsilon^3 \mu^3}{2(1 - \epsilon\mu)} = \Phi, \end{aligned}$$

for  $\epsilon\mu < 1$ . Since  $Z$  is of order  $\epsilon$ ,  $\Phi$  will ordinarily be small compared with  $Z$ ; therefore consider the nature of the distribution of  $Z$ .

If we write  $Z = a_1 x_1 + \dots + a_p x_p$ , then, since the  $x$ 's are independently distributed with rectangular distributions, it is easily shown that  $\mu_2 = \frac{\epsilon^2}{3} \sum a_i^2$ ,  $\alpha_3 = 0$ ,  $\alpha_4 = 3 - \frac{8}{3} \sum a_i^4 / (\sum a_i^2)^2$ . If the  $a_i$  are approximately equal in magnitude, then  $\alpha_4$  is approximately equal to  $3 - 1/p$ . But from (15)  $p \geq \frac{1}{2}(n - k)^2 + \dots + \frac{1}{2}(n - 1)^2$ , which is sufficiently large for determinants employing Chio's method to justify the assumption that  $Z$  is approximately normally distributed. Setting  $L = h_k^{n-k-2} \dots h_1^{n-3}$ ,

$$\begin{aligned} \mu_2 &= \frac{L^2 \epsilon^2}{3} \left[ \left( \sum_{i=1}^n \Delta_{ii}^2 + 4 \sum_{i < j} \Delta_{ij}^2 \right) + \dots + h_k^2 \dots h_2^2 \left( \sum_{i=2}^n \Delta_{ii}^2 + 4 \sum_{i < j} \Delta_{ij}^2 \right) \right] \\ &\leq \frac{2\epsilon^2}{3} [(n - k)^2 + \dots + (n - 1)^2 - \frac{1}{2}\{(n - k) + \dots + (n - 1)^2\}] \\ &\leq \frac{2\epsilon^2}{3} \left[ (n - k)^2 + \dots + (n - 1)^2 - \frac{k}{4}(2n - k - 1) \right] = \Psi^2. \end{aligned}$$

Hence, the probability is  $>.95$  that  $|Z| < 2\Psi$ . Since  $|I - Z| \leq \Phi$ , the probability is  $>.95$  that  $|I| < 2\Psi + \Phi$ ; and therefore the probability is  $>.95$  that

$$(16) \quad |J| < \frac{2\Psi + \Phi}{C}.$$

This inequality will usually give a smaller bound for  $|J|$  than (8). However, when  $\Delta$  is small the  $H$ 's may be small, with the result that  $C$  will be small and (16) may not give a satisfactory bound for  $|J|$ . In such cases the bound given by (8) may not prove satisfactory either.

**4. Example.** Consider a correlation determinant of order 7 in which the elements are accurate to 4 decimal places. If Chio's reduction method is applied until a 2 rowed determinant is obtained, then  $n = 7$ ,  $k = 5$ ,  $\epsilon = .00005$ ,  $m = 90$ ,  $\mu = 176$ ,  $\Psi = .00005\sqrt{160/3}$ , and we obtain from (8) that

$$|J| < \left(\frac{H}{\bar{H}}\right).0045 + \left(\frac{H}{\bar{H}}\right)^2 .00001 + \left(\frac{H}{\bar{H}}\right)^3 \frac{.00000005}{1 - .0045 H/\bar{H}}$$

where  $\bar{H}/H$  is obtained from calculations involved in evaluating the determinant. From (16) we obtain that the probability is  $>.95$  that

$$|J| < \frac{.0008}{C}.$$

The relative advantage of the second inequality over the first depends on the size of the pivotal elements, as does the usefulness of either inequality.

UNIVERSITY OF CALIFORNIA AT LOS ANGELES



# THE CUMULATIVE NUMBERS AND THEIR POLYNOMIALS

BY P. S. DWYER

In a recent paper [1] the author has shown how the moments of a distribution can be obtained from the last entries of cumulative columns with the use of multiplication by certain numbers. These numbers may be called "cumulative numbers." It is the aim of this paper to show how these numbers can be obtained from the expansion of  $x^n$  in terms of factorials of the  $s$ -th order and to demonstrate properties of the polynomials of which these numbers are the coefficients.

TABLE 1

*Successive Frequency Cumulations*

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$X$	$x$	$f_x$	$C^1$	$C^2$	$C^3$	$C^4$	$C^5$
$a + 6$	6	64	64	64	64	64	64
$a + 5$	5	192	256	320	384	448	512
$a + 4$	4	240	496	816	1200	1648	2160
$a + 3$	3	160	656	1472	2672	4320	6480
$a + 2$	2	60	716	2188	4860	9180	15660
$a + 1$	1	12	728	2916	7776	16956	32616
$a$	0	1	729	3645	11421	28377	60993

1. **The values  $C_i^j(u_x)$ .** We use the notation  $C_i^j(u_x)$  of the previous paper [1,289] to express the columnar cumulated entries. The  $j$  indicates the order of the cumulation while the  $i$  indicates the number of the term, counting from the bottom of the column. Thus in Table I, which presents the cumulations of a frequency distribution used in the previous paper [1,289],  $C_1^1 = 729$ ;  $C_1^2 = 3645$ ;  $C_2^2 = 2916$ ;  $\dots$ ,  $C_4^5 = 6480$ , etc. Now if  $k + 1$  values of  $x$  are spaced at unit distances and if the smallest value of  $x$  is 0, it can be shown that

$$C_1^1 = \sum u_x; \quad C_1^2 = \sum (x + 1)u_x; \quad C_2^2 = \sum xu_x; \quad C_1^3 = \sum \frac{(x + 2)(x + 1)}{2!} u_x;$$

$$C_2^3 = \sum_0^k \frac{(x + 1)x}{2!} u_x; \quad C_3^3 = \sum_0^k \frac{x(x - 1)}{2!} u_x$$

and, in general,  $j > 0$  and  $j + 1 \geq i$ ,

$$(1) \quad C_i^{j+1} = \sum_{x=0}^k \frac{(x + j + 1 - i)^{(j)}}{j!} u_x.$$

Similarly if  $k$  values of  $x$  are spaced at unit distances and if the smallest value of  $x$  is 1, it can be shown that

$$C_1^1 = \sum_1^k u_x; \quad C_1^2 = \sum_1^k x u_x; \quad C_2^2 = \sum_1^k (x-1) u_x; \quad C_1^3 = \sum_1^k \frac{(x+1)x}{2!} u_x; \\ C_2^3 = \sum_1^k \frac{x(x-1)}{2!} u_x; \quad C_3^3 = \sum_1^k \frac{(x-1)(x-2)}{2!} u_x$$

and, in general,  $j > 0$  and  $j+1 \geq i$ ,

$$(2) \quad C_i^{j+1} = \sum_{x=1}^k \frac{(x+j-i)^{(j)}}{j!} u_x.$$

It is to be noted that the coefficients of  $u_x$  in (2) could be obtained from the coefficients of  $u_x$  in (1) by the substitution  $x+1 = x'$ .

**2. The powers in terms of factorials of the  $s$ -th order.** If the  $s$ -th powers can be expressed in terms of factorials of the  $s$ -th order (factorials having  $s$  factors) then the moments can be expressed in terms of the cumulations. For example

$$x^2 = \frac{(x+1)x + x(x-1)}{2} \text{ so, from (1)}$$

$$\sum_0^k x^2 f_x = \sum_0^k \frac{(x+1)^{(2)}}{2!} f_x + \sum_0^k \frac{x^{(2)}}{2!} f_x = C_2^2 + C_3^2.$$

And since

$$x^3 = \frac{(x+2)^{(3)} + 4(x+1)^{(3)} + x^{(3)}}{3!}, \text{ we have}$$

$$\sum_0^k x^3 f_x = \sum_0^k \frac{(x+2)^{(3)}}{3!} f_x + 4 \sum_0^k \frac{(x+1)^{(3)}}{3!} f_x + \sum_0^k \frac{x^{(3)}}{3!} f_x = C_2^4 + 4C_3^4 + C_4^4.$$

In general if

$$(3) \quad x^s = \frac{A_{s1}(x+s-1)^{(s)} + A_{s2}(x+s-2)^{(s)} + \dots + A_{sj}(x+s-j)^{(s)} + \dots + A_{ss}x^{(s)}}{s!},$$

then

$$(4) \quad \sum_0^k x^s f_x = A_{s1}C_2^{s+1} + A_{s2}C_3^{s+1} + \dots + A_{sj}C_{j+1}^{s+1} + \dots + A_{ss}C_{s+1}^{s+1},$$

while if the smallest value of  $x$  is 1, we have

$$(5) \quad \sum_1^k x^s f_x = A_{s1}C_1^{s+1} + A_{s2}C_2^{s+1} + \dots + A_{sj}C_j^{s+1} + \dots + A_{ss}C_s^{s+1}.$$

These quantities,  $A_{sj}$ , in (4) and (5) are simply the coefficients of certain factorials of the  $s$ -th order in the expansion of  $x^s s!$ .

These numbers, for small values of  $s$ , are easily obtained. It is possible to use the table and a recursion formula of a previous paper [1,294-295] for larger values of  $s$ . It is also possible to obtain these values, without involving cumulative theory, from (3) above.

While doing this we make a more general approach by expanding  $(a + x)^s$  in terms of these same factorials with the coefficients now functions of  $a$ . This is possible if we add an additional term,  $A_{s0}(x + s)^{(s)}$ , to the numerator of the right hand side of (3). We have then

$$(6) \quad (a + x)^s = \frac{A_{s0}(x + s)^{(s)} + A_{s1}(x + s - 1)^{(s)} + \dots + A_{sj}(x + s - j)^{(s)} + \dots + A_{ss}x^{(s)}}{s!}$$

The determination of the values  $A_{sj}$  can be accomplished by purely algebraic means by successive substitution of  $x = 0, 1, 2, \dots, s$ . In this way we obtain  $s + 1$  equations in  $s + 1$  unknowns. For example when  $s = 2$

$$(a + x)^2 = \frac{A_{20}(x + 2)^{(2)} + A_{21}(x + 1)^{(2)} + A_{22}x^{(2)}}{2!}$$

so that when  $x = 0, 1, 2$ , we have

$$a^2 = A_{20}; (a + 1)^2 = 3A_{20} + A_{21}; (a + 2)^2 = 6A_{20} + 3A_{21} + A_{22}.$$

The solution is  $A_{20} = a^2$ ;  $A_{21} = 2ab + 1$ ;  $A_{22} = b^2$  where  $b = 1 - a$ . It follows that

$$(a + x)^2 = a^2 \frac{(x + 2)^{(2)}}{2!} + (2ab + 1) \frac{(x + 1)^{(2)}}{2!} + b^2 \frac{x^{(2)}}{2!} \text{ and hence that}$$

$$\sum_0^k (a + x)^2 f_x = a^2 C_1^3 + (2ab + 1) C_2^3 + b^2 C_3^3,$$

as indicated in the previous paper [1,293].

When  $a = 0$ , then  $b = 1$  and we have

$$\sum x^2 f_x = C_1^3 + C_3^3 \text{ while when } a = 1, b = 0 \text{ and the right hand side becomes } C_1^3 + C_2^3.$$

It follows that the general cumulative numbers might also be defined as the solutions of the  $s + 1$  equations in the  $s + 1$  unknowns obtained by placing  $x = 0, 1, 2, \dots, s$  in (6).

**3. The evaluation of the cumulative numbers.** Formal algebraic methods of evaluating equations (6) are somewhat tedious so we use finite difference theory to aid in finding the solution. As in the previous paper [1] we use the notation

$$\nabla v_x = v_x - v_{x-1} \text{ and } v_x = \begin{cases} v_x & \text{when } a \leq x \leq a + k \\ 0 & \text{otherwise} \end{cases}. \text{ We then write, from (6)}$$

$$(7) \quad \underline{s!(a+x)^s} = A_{s0}\underline{(x+s)^{(s)}} + A_{s1}\underline{(x+s-1)^{(s)}} \\ + \dots + A_{sj}\underline{(x+s-j)^{(s)}} + \dots + A_{ss}\underline{x^{(s)}}.$$

We note further that  $\nabla^{s+1}(x+r)^{(s)} = \begin{cases} s! & \text{when } r=0 \\ 0 & \text{when } r \neq 0 \end{cases}$ . We have then

$$(8) \quad \nabla^{s+1}(a+j)^s = A_{sj}.$$

It has been shown in the previous paper [1,292] that

$$(9) \quad \nabla^{s+1}(a+j)^s = \sum_{i=0}^j (-1)^i \binom{s+1}{i} (a+j-i)^s$$

and it appears that the cumulative numbers could be defined by (9). A useful recursion formula has been derived from (9)

$$(10) \quad \nabla^{s+1}(a+x)^s = (a+x)\nabla^s(a+x)^{s-1} + (s+1-a-x)\nabla^s(a+x-1)^{s-1}.$$

**4. The cumulative polynomials.** We define the cumulative polynomials to be the polynomials obtained by using the cumulative numbers as coefficients. Thus when  $a=0$ ,

$$P_1 = y; \quad P_2 = y + y^2; \quad P_3 = y + 4y^2 + y^3; \quad P_4 = y + 11y^2 + 11y^3 + y^4; \text{ etc.}$$

It is possible to derive a recursion formula for these polynomials. We use (10) with  $s$  replaced by  $s+1$  and  $a=0$  and get

$$(11) \quad P_{s+1} = \Sigma \nabla^{s+2}(x)^{s+1} y^x = \Sigma x \nabla^{s+1}(x)^s y^x + \Sigma (s+2-x) \nabla^{s+1}(x-1)^s y^x,$$

which becomes, after some manipulation,

$$(12) \quad P_{s+1} = (1-y) \Sigma x \nabla^{s+1}(x)^s y^x + (s+1)y P_s.$$

To illustrate we get  $P_4$  from  $P_3 = y + 4y^2 + y^3$ . Now  $\Sigma x \nabla^4(x)^3 y^x = y + 8y^2 + 3y^3$  and  $P_4 = (1-y)(y + 8y^2 + 3y^3) + 4y(y + 4y^2 + y^3) = y + 11y^2 + 4y^3 + y^4$ . The recursion formula (12) can be expressed also in the form of a differential equation, since  $P'_s = \frac{d}{dy}(P_s) = \Sigma x \nabla^{s+1}(x)^s y^{x-1}$ , as

$$(13) \quad P_{s+1} = y[(1-y)P'_s + (s+1)P_s].$$

It can be shown more generally that for any  $a$

$$P_{a,0} = 1; \quad P_{a,1} = a + by; \quad P_{a,2} = a^2 + (2ab + 1)y + b^2y^2, \text{ etc. with}$$

$$(14) \quad P_{a,s+1} = y(1-y)P'_{a,s} + [a(1-y) + (s+1)y]P_{a,s}$$

as the recursion formula.

**5. The numerator coefficients in successive derivatives of the logistic function.** Lotka has recently exhibited the coefficients of the numerator terms of suc-

cessive derivatives of the logistic function [2, 160]. These appear to be, aside from sign, the same as the cumulative numbers when  $a = 0$ . It is shown in this section that these numbers are the cumulative numbers. The scheme is generalized to include the numerator coefficients of the derivatives of a more general function involving the parameter  $a$ .

Lotka used the function  $\Phi_0 = \frac{1}{1 + e^{rt}}$  and obtained  $\Phi_1 = \frac{re^{rt}}{(1 + e^{rt})^2}$ ,  $\Phi_2 = \frac{r^2 e^{rt}(1 - e^{rt})}{(1 + e^{rt})^3}$ , etc. The numerical coefficients are the same if  $r = 1$  so we might as well use  $\Phi_0 = \frac{1}{1 + e^x}$ . A more general function is the two parameter function

$$(15) \quad \Phi_{a,c} = \frac{e^{ax}}{1 + ce^x}.$$

Let successive derivatives with respect to  $x$  be indicated by  $\Phi_{a,c,1}$ ;  $\Phi_{a,c,2}$ ;  $\Phi_{a,c,3}$ ; etc. Then

$$\begin{aligned} \Phi_{a,c,1} &= \frac{e^{ax}[a + c(1 - a)e^x]}{(1 + ce^x)^2}, \\ \Phi_{a,c,2} &= \frac{e^{ax}[a^2 + (-2a^2 + 2a + 1)ce^x + (1 - a)^2 c^2 e^{2x}]}{(1 + ce^x)^3} \end{aligned}$$

In general,

$$\Phi_{a,c,s} = \frac{e^{ax} Q_{a,c,s}}{(1 + ce^x)^{s+1}} = e^{ax} Q_{a,c,s} (1 + ce^x)^{-s-1}$$

so that

$$\Phi_{a,c,s+1} = \frac{e^{ax}\{(1 + ce^x)[aQ_{a,c,s} + Q'_{a,c,s}] - (s + 1)ce^x Q_{a,c,s}\}}{(1 + ce^x)^{s+2}}$$

and

$$(16) \quad Q_{a,c,s+1} = (1 + ce^x)[aQ_{a,c,s} + Q'_{a,c,s}] - (s + 1)ce^x Q_{a,c,s}.$$

The  $Q$  functions can be changed to polynomials with the substitution  $e^x = y$ . Then derivatives are taken with respect to  $y$  and

$$(17) \quad P_{a,c,s+1} = (1 + cy)[aP_{a,c,s} + yP'_{a,c,s}] - (s + 1)cyP_{a,c,s}.$$

When  $c = -1$ , this becomes formula (14) and since  $P_{a,0} = 1$ , it follows that the numbers of the present section are generalized cumulative numbers. When  $c = 1$  and  $a = 0$  we have the numbers found by Lotka.

It can be shown, further, that the  $c$  coefficient of  $y^j$  is  $c^j$ . It follows that the absolute values of the coefficients, when  $c = 1$  and when  $c = -1$ , are the same.

**6. Formulas for  $\Sigma x^s$ .** A formula for the sums of the  $s$ -th powers of the integers from 1 to  $k$  is obtained by summing (3). We get

$$(18) \quad \sum_1^k x^s = A_{s1} \sum_1^k \frac{(x+s-1)^{(s)}}{s!} + \dots \\ + A_{sj} \sum_1^k \frac{(x+s-j)^{(s)}}{s!} + \dots + A_{sk} \sum_1^k \frac{x^{(s)}}{s!}$$

from which

$$(19) \quad \sum_1^k x^s = A_{s1} \frac{(k+s-1)^{(s+1)}}{(s+1)!} + \dots \\ + A_{sj} \frac{(k+s-j)^{(s+1)}}{(s+1)!} + \dots + A_{sk} \frac{k^{(s+1)}}{(s+1)!},$$

or

$$(20) \quad \sum_1^k x^s = \sum_{j=1}^s A_{sj} \frac{(k+s-j)^{(s+1)}}{(s+1)!} = \frac{1}{(s+1)!} \sum_{j=1}^s (k+s-j)^{(s+1)} \nabla^{s+1} \underline{(j)^s} \\ = \sum_{j=1}^s C_{j+1}^{s+1} (1) \nabla^{s+1} \underline{(j)^s}.$$

For example

$$\sum_1^k x^2 = \frac{(k+2)^{(3)} + (k+1)^{(3)}}{3!} = \frac{k(k+1)(2k+1)}{6}, \\ \sum_1^k x^3 = \frac{(k+3)^{(4)} + 4(k+2)^{(4)} + (k+1)^{(4)}}{4!} = \frac{k^2(k+1)^2}{4}.$$

More generally the values of  $\sum_a^{a+k} x^s$  can be evaluated by

$$(21) \quad \sum_a^{a+k} x^s = \frac{1}{(s+1)!} \sum_{j=0}^s (k+s-j)^{(s+1)} \nabla^{s+1} \underline{(a+j)^s} = \sum_{j=0}^s C_{j+1}^{s+1} (1) \nabla^{s+1} \underline{(a+j)^s}.$$

**7. Summary.** It is shown how the cumulative numbers and the cumulative polynomials may be obtained in a variety of ways. Of special interest is the fact that the cumulative numbers can be obtained by expanding powers in terms of factorials and hence they might be called factorial coefficients of a kind. It is also possible, though it is not within the scope of this paper, to establish interesting relations between the cumulative numbers and the multinomial coefficients, the usual factorial coefficients, the difference of 0, etc.

#### REFERENCES

- [1]. P. S. DWYER. "The Computation of Moments with the use of Cumulative Totals." *Annals of Math. Stat.* Vol. IX, no. 4, Dec. 1938, pp. 288-304. A more extensive bibliography is available here.
- [2]. A. J. LOTKA. "An Integral Equation in Population Analysis." *Annals of Math. Stat.* Vol. X, no. 2, June 1939, pp. 144-161.

UNIVERSITY OF MICHIGAN  
ANN ARBOR, MICHIGAN

# ENUMERATION AND CONSTRUCTION OF BALANCED INCOMPLETE BLOCK CONFIGURATIONS<sup>1</sup>

BY GERTRUDE M. COX

1. **Introduction.** One of the general problems of experimental design is to avoid extraneous effects in making desired comparisons. The method employed is to use experimental materials as nearly homogeneous as possible. Such materials, however, are seldom available in large quantities. On the contrary, field soils vary in fertility from block to block, animals vary with both litter and sex, and leaves on one young plant differ from those on another. Differences between blocks, between litters and sex, and between plants, being irrelevant to the comparisons usually contemplated, must be avoided.

When the number of treatments to be compared is small, well known methods of design, such as the Latin square or randomized complete block, are available and efficient. As the number of treatments increases, however, these designs tend to become less efficient through failure to eliminate heterogeneity. Furthermore, they become cumbersome, the Latin square design requiring replicates equal in number to the treatments and the complete block design providing that each treatment occur in every block. (Blocks are defined as an assemblage of experimental units chosen to be as nearly alike as possible.)

Because of such limitations, several modifications of the complete block design have been devised. These new designs all have the common characteristic that the experimental material is divided into groups or blocks containing fewer units than the number of treatments to be compared. These more homogeneous small blocks are referred to as incomplete blocks.

It is desirable to have all comparisons between pairs of treatments made with equal accuracy. This requires of the design that every pair of treatments occur in the same block an equal number of times. Such a design is referred to as balanced. Balanced incomplete block designs can be arranged (for any given number of treatments) only for certain combinations of block size and number of replications.<sup>2</sup>

The construction of balanced incomplete block designs is mathematically a part of the theory of configurations. A configuration is an assemblage of elements into sets, each element occurring in the same number of sets, and each

<sup>1</sup> A revision of an expository paper presented under a different title at a joint meeting of the Institute of Mathematical Statistics and Biometric Section of the American Statistical Association, December 27, 1939.

<sup>2</sup> Numerous additional designs are available in the partially balanced incomplete blocks [3].

set containing the same number of elements. The configurations to be considered here are the complete configurations, i.e., those in which each element occurs an equal number of times in the same set with every other element. It would be useful to know, (a) what configurations (within the useful range) exist. (b) how these configurations may be constructed.

The typical requirement of the experimenter is this: "I wish to test  $t$  treatments and can use blocks of size  $k$  ( $t > k$ ). I should like a design which will involve as little experimental material as feasible." The designer must then determine what configuration of  $t$  elements in sets of  $k$  will satisfy the incidence relation that each pair of elements occur together in a set an equal number of times, and for which the total number of sets is a minimum. There are still many configurations which the experimenter needs but which have not as yet been constructed.

In order better to explain the construction of these balanced incomplete block designs, it is essential to specify the underlying combinatorial problems. A configuration satisfying the condition of balance can be obtained by writing down all possible combinations,  $b$ , of the  $t$  elements taken  $k$  at a time,

$$b = {}_tC_k = \frac{t!}{k!(t-k)!}.$$

The simplest example is that in which each set contains only two elements and all possible combinations of the  $t$  elements, taken in pairs, appear in the different sets. This series of pairs can be written out by the experimenter, and the method of analysis is given by Yates [20].

Let us take another example; given six elements to be taken three at a time,

$$b = {}_6C_3 = \frac{6!}{3!3!} = 20.$$

The 20 combinations are,

	134	146	236	345
124	135	156	245	346
125	136	234	246	356
126	145	235	256	456.

Such unreduced designs are not necessarily economical or feasible in experimental work. It is often desirable to find some less extensive configuration. In this example half of the combinations, either those in italics or the other half, fulfill the restriction that every element occur with every other element in the same number of sets. Each pair of elements occurs twice in either group of sets. Thus, a balanced incomplete block design can be based on either half of the 20 sets as well as on all 20.

**2. Combinatorial methods.** Combinatorial considerations of a simple nature enable us to set up necessary conditions which balanced designs must satisfy.



We have  $t$  elements arranged in  $b$  sets of  $k$  elements each; each element occurs in  $r$  sets, and each pair of elements occurs together in a set exactly  $\lambda$  times. Then we must have

$$tr = bk, \quad r(k - 1) = \lambda(t - 1).$$

The first of these equations expresses the fact that the total number of plots must be equal both to the product of elements by replications and to the product of sets by number of elements per set; the second, that the number of pairs into which a given element enters must equal  $\lambda$  times the remaining number of elements.

It is convenient to write

$$r = \frac{\lambda(t - 1)}{k - 1}, \quad b = \frac{\lambda t(t - 1)}{k(k - 1)}.$$

Since the numbers  $t, b, r, k, \lambda$  must be integers, it is easy to obtain lower limits for any three in terms of the other two.

To give a general classification, the configurations have been divided into classes according to the value of  $\lambda$ . Because of the practical limitations in experimentation, table I has been expanded only to include  $\lambda = 6$  and the  $k$  values from 1-14. It may be well to call attention to the fact that duplications occur in the different classes of table I. For instance in the class,  $\lambda = 1$ , for  $k = 6, t = 15m + 1$ , and  $m = 1$ , then  $b = 8$ , and  $r = 3$ . In order to construct a design, the following condition is necessary;  $r \geq k$  and therefore  $b \geq t$ . In this example, the condition is met if  $b, r$  and  $\lambda$  are multiplied by 2, the resulting design is  $t = 16, b = 16, r = 6, k = 6$  and  $\lambda = 2$ . This configuration is a duplicate of the design in the class,  $\lambda = 2$ , for  $k = 6$  and  $m = 1$ . In many of the configurations where  $\lambda$  is 3, 4, 5, or 6, a common factor can be cancelled from  $b, r$  and  $\lambda$  giving a design listed in the classes,  $\lambda = 1, 2$  or 3.

It should be emphasized that the conditions under which table I was derived are necessary, but not sufficient, for the existence of a complete configuration. For example, consider the following configurations which satisfy the necessary conditions for a design.

Sub class (table I)	$m$	$t$	$b$	$r$	$k$	$\lambda$
$10m + 5$	1	15	21	7	5	2
$21m + 1$	1	22	22	7	7	2
$15m + 6$	2	36	42	7	6	1
$42m + 1$	1	43	43	7	7	1
$45m + 10$	2	100	110	11	10	1
$110m + 1$	1	111	111	11	11	1.

No configurations of the above specification can actually be constructed.

A selected group of configurations from table I is given in table II. Only those configurations whose  $k, r$  and  $\lambda$  lie within practical limits, and whose

TABLE I  
Classes of Configurations

Class

$k$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$
	$t$	$t$	$t$	$t$	$t$	$t$
3	$6m + 1$	$3m + 1$	$2m + 1$	$3m + 1$		
4	$12m + 1$	$6m + 1$	$4m + 1$	$5m + 1$	$4m + 1$	
5	$20m + 1$	$10m + 1$	$20m + 1$	$5m + 1$	$3m + 1$	$5m + 1$
6	$15m + 1$	$15m + 1$	$10m + 1$	$15m + 1$	$3m + 1$	$7m + 1$
7	$42m + 1$	$21m + 1$	$14m + 1$	$21m + 1$	$42m + 1$	$28m + 1$
8	$56m + 1$	$28m + 1$	$56m + 1$	$14m + 1$	$56m + 1$	$12m + 1$
9	$72m + 1$	$36m + 1$	$24m + 1$	$18m + 1$	$72m + 1$	$15m + 1$
10	$45m + 1$	$45m + 1$	$30m + 1$	$45m + 1$	$18m + 1$	$55m + 1$
11	$110m + 1$	$55m + 1$	$110m + 1$	$55m + 1$	$22m + 1$	$22m + 1$
12	$132m + 1$	$66m + 1$	$44m + 1$	$33m + 1$	$132m + 1$	$26m + 1$
13	$156m + 1$	$78m + 1$	$52m + 1$	$39m + 1$	$156m + 1$	$91m + 1$
14	$91m + 1$	$91m + 1$	$182m + 1$	$91m + 1$	$91m + 1$	
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

$m = 1, 2, 3, \dots$

existence has not been disproved, have been included. The practical limits of  $k$ ,  $r$  and  $\lambda$ , of course, are dependent upon the conditions surrounding the experiment. We have chosen to keep  $k$  within the range 3 to 10 except for a few special configurations in which  $t$  is greater than 100, in which cases  $k$  was allowed to equal 11-14. Also  $r$  has been kept within a similar limited range. (Those configurations in table II, with an asterisk preceding  $t$ , have not been constructed.)

The above limitations upon  $k$  and  $r$  give a small, selected group of configurations. However, many others either have been constructed or are known to exist. For balanced incomplete block designs, Yates [20] gives the lower limits of  $r$  for  $t$  from 4 to 25 and  $k$  from 2 to 12 but not greater than  $\frac{1}{2}t$ . Fisher and Yates [8] have tabulated the configurations which are known to exist having ten or less replications including all arithmetically possible configurations the existence of which has not been disproved.

Even if the existence of a configuration has not been disproved, there still remains the difficult problem of writing out the elements which are to appear in each set. Some discussion of the structure of such configurations is presented by Fisher and Yates [8] by Yates [20, 21] by Goulden [9, 10] and by Bose [4]. Additional descriptions are to follow.

While a search of the literature revealed a number of constructed configurations, yet the general theory of their formation has received relatively little consideration. The question of combinations related to the theory of configurations which is of interest here was first set forth by Kirkman [11] in 1847. He states the problem thus: "If  $Q_x$  denote the greatest number of triads that can be formed with  $x$  symbols, so that no duad shall be twice employed, then

$$3Q_x = x(x-1)/2 - V_x$$

if for  $V_x$  we put 0, when  $x = 6m + 1$  or  $6m + 3$ ." This gives the formula for  $b$  which was given earlier in this article. Put  $x = t$  and  $V_x = 0$

$$b = Q_x = \frac{t(t-1)}{3 \cdot 2} = \frac{t(t-1)}{k(k-1)}.$$

Besides the theory connected with these combinatorial problems, considerable information related to the construction of the configurations has been found in the literature on finite projective geometry, especially the geometry which applies to the theory of groups.

An extensive discussion of the  $\lambda = 1$  class of configurations (as listed in table I) can be found in the literature. The theory of the formation of the configurations for the sub-class  $t = 6m + 3$  has been summarized by Ball [1]. This is the Kirkman "school-girl problem" for which Eckenstein [7] lists 48 papers and 5 books written during the years 1847-1911 dealing with this subject. The problem was first published in the *Lady's and Gentleman's Diary* for 1850 [12]. It is usually stated that "a schoolmistress was in the habit of taking her girls for a daily walk. The girls were fifteen in number, and were arranged in five rows of three each, so that each girl might have two companions. The problem

is to dispose of them so that for seven consecutive days no girl will walk with any of her school-fellows in any triplet more than once." For this particular subclass ( $t = 6m + 3$ ,  $k = 3$ ), this type of configuration has been shown to exist

TABLE II  
*Selected Group of Configurations*  
(Balanced Incomplete Block Designs)

$t$	$b$	$r$	$k$	$\lambda$	$t$	$b$	$r$	$k$	$\lambda$
7	7	3	3	1 Y.S. <sup>1</sup>	*25	50	8	4	1
7	7	4	4	2	25	30	6	5	1
8	14	7	4	3	25	15 + 15	3	5	1 L.S.
9	12	4	3	1	*25	25	9	9	3
9	6 + 6	2	3	1 L.S. <sup>2</sup>	28	63	9	4	1
9	18	8	4	3	28	36	9	7	2
9	18	10	5	5	*29	29	8	8	2
9	12	8	6	5	31	31	6	6	1 Y.S.
10	30	9	3	2	*31	31	10	10	3
10	15	6	4	2	*36	45	10	8	2
10	18	9	5	4	37	37	9	9	2
10	15	9	6	5	*41	82	10	5	1
11	11	5	5	2	*46	69	9	6	1
11	11	6	6	3	*46	46	10	10	2
13	26	6	3	1	49	56	8	7	1
13	13	4	4	1 Y.S.	49	28 + 28	4	7	1 L.S.
13	13	9	9	6	*51	85	10	6	1
15	35	7	3	1	57	57	8	8	1 Y.S.
15	15	7	7	3	64	72	9	8	1
15	15	8	8	4	64	72 + 72	9	8	2 L.S.
16	20	5	4	1	73	73	9	9	1 Y.S.
16	20 + 20	5	4	2 L.S.	81	90	10	9	1
16	16	6	6	2	81	45 + 45	5	9	1 L.S.
16	16	10	10	6	91	91	10	10	1 Y.S.
19	57	9	3	1	121	132	12	11	1
19	19	9	9	4	121	66 + 66	6	11	1 L.S.
19	19	10	10	5	133	133	12	12	1 Y.S.
21	70	10	3	1	169	182	14	13	1
21	21	5	5	1 Y.S.	169	91 + 91	7	13	1 L.S.
*21	28	8	6	2	183	183	14	14	1 Y.S.
*21	30	10	7	3					

\* Have not been constructed.

<sup>1</sup> Youden squares.

<sup>2</sup> Lattice squares.

for every possible value of  $t$ . Most of the solutions were worked by H. E. Dudeney and O. Eckenstein. They are given by Ball [1] for all  $t$ 's less than 100, that is, for  $t = 9, 15, 21, 27, 33, 39, 45, 51, 57, 63, 69, 75, 81, 87, 93$  and  $99$ . Ball describes several methods of constructing such configurations, as cycles, combinations of cycles, scalene triangles inscribed in the circle, focal and analyti-

cal methods. As an illustration of the school-girl problem, the construction of the configuration for  $t = 9$ ,  $b = 12$ ,  $r = 4$ ,  $k = 3$  and  $\lambda = 1$  will be shown. Scalene triangles are inscribed in a circle with certain specifications (to be fulfilled) giving the three sets of triplets for the first day as follows,

Set	Group I		
(1)	$k$	1	5
(2)	3	4	6
(3)	7	8	2.

By rotation or by cyclic substitution the other three groups are secured:

Set	Group II			Group III			Group IV				
(4)	$k$	2	6	(7)	$k$	3	7	(10)	$k$	4	8
(5)	4	5	7	(8)	5	6	8	(11)	6	7	1
(6)	8	1	3,	(9)	1	2	4,	(12)	2	3	5.

Then placing  $k = 9$ , we have the configuration for  $t = 9$ ,  $b = 12$ , and  $r = 4$ . Note that in the school-girl problem the sets are grouped into complete replications of the elements. This problem of 9 girls taken 3 at a time has been subjected to an exhaustive examination. There are 840 arrangements but only one fundamental solution. In the case of 15 girls, the number of fundamental solutions according to Mulden [14] and Cole [6], is seven. Ball mentions the Kirkman problem in quartets which is the sub-class  $t = 12m + 4$ , for  $k = 4$ . He states that this has been solved for cases where  $m$  does not exceed 49. He also states, "I conjecture that similar methods are applicable to corresponding problems about quintets, sextets, etc."

Before leaving the school-girl problem, an illustration will be given of  $t = 28$ ,  $b = 63$ ,  $r = 9$ ,  $k = 4$  and  $\lambda = 1$ . The following framework was set up by Dr. C. P. Winsor using suggestions from Netto [15].

$a_1$	$a_8$	$b_3$	$b_6$
$a_2$	$a_7$	$b_1$	$b_8$
$a_3$	$a_6$	$c_4$	$c_6$
$a_4$	$a_5$	$c_1$	$c_8$
$b_2$	$b_7$	$c_3$	$c_5$
$b_4$	$b_5$	$c_2$	$c_7$ .

$a$ ,  $b$  and  $c$  each have every internal difference once and only once; and each pair  $a$ - $b$ ,  $a$ - $c$  and  $b$ - $c$  must have every external difference once and only once. The nine groups are given in table III. The cyclic substitution is within three sets,  $a$ ,  $b$  and  $c$ . That is,



$l = (t) =$  number of points in the geometry,  
 $m = (b) =$  number of lines,  
 $\lambda = (k) =$  number of points,  
 $\mu = (r) =$  number of lines on a point.

The series of finite projective geometries  $PG(\kappa, p^n)$  for  $\kappa > 1$  furnishes a certain infinite class of these tactical configurations. The following list gives those which have been incorporated in the list (table II) of useful balanced incomplete block designs.

*Two dimensional space,  $PG(2, p^n)$*

$p^n$	$l(t)$	$m(b)$	$\lambda(k)$	$\mu(r)$
2	7	7	3	3
3	13	13	4	4
$2^2$	21	21	5	5
5	31	31	6	6
7	57	57	8	8
$2^3$	73	73	9	9
$3^2$	91	91	10	10
11	133	133	12	12
13	183	183	14	14.

*Three dimensional space,  $PG(3, p^n)$*

$p^n$	$l$	$m$	$\lambda$	$\mu$
2	15	35	7	3.

From the Euclidean geometry  $EG(\kappa, p^n)$  for  $\kappa > 1$  other tactical configurations can be constructed. These are formed from the  $PG(\kappa, p^n)$  by omitting a given line from the two dimensional space and a plane from the three dimensional space configurations. Some of the resulting designs are:

*Two dimensional space,  $EG(2, p^n)$*

$p^n$	$l$	$m$	$\lambda$	$\mu$
2	4	6	3	2
3	9	12	4	3
$2^2$	16	20	5	4
5	25	30	6	5
7	49	56	8	7
$2^3$	64	72	9	8
$3^2$	81	90	10	9
11	121	132	12	11
13	169	182	14	13.

Methods are available for constructing the two dimensional space  $PG(\kappa, p^n)$  and the corresponding  $EG(\kappa, p^n)$  configurations where  $p$  is a prime number. This being true, we can also construct the completely orthogonalized squares from the  $EG(\kappa, p^n)$  geometry. The reverse situation in which these configurations are constructed by using the completely orthogonalized squares is to be illustrated. These squares consist of superimposed Latin squares, fulfilling the condition that each number from the second Latin square occurs once and only once with each number in the first Latin square. As an example take the two Latin squares:

Latin Square I			Latin Square II		
1	2	3	1	3	2
2	3	1	2	1	3
3	1	2	3	2	1

Superimpose square II upon square I to get the completely orthogonalized  $3 \times 3$  square,

11	23	32
22	31	13
33	12	21

The first number in each cell is a value from square I; the second number in each cell is from square II. Note that the numbers in the second place in each cell occur once and only once with each of the first numbers, that is 1-1, 1-3, and 1-2. The completely orthogonalized squares have been proven to exist for all prime numbers and for powers of prime numbers. The solution of this problem was secured independently by Bose [2] and by Stevens [18]. Those of sides  $2, 2^2, 2^3, 2^4, 2^5, 2^6, 3, 3^2, 3^3, 3^4, 5, 5^2, 5^3, 7, 7^2, 11$  and  $13$  have been given.

The completely orthogonalized  $3 \times 3$  square may be used to construct

11	23	4	32	7
22	31	5	13	8
33	12	6	21	9

a balanced incomplete block design. The italic numbers, which follow the cell numbers, designate the 9 elements which are to be arranged in four groups of three sets. Group I is formed by placing the elements from each row into separate sets; in group II the elements from the three columns are placed in three sets; in group III the first set (7) consists of the elements which follow 1 in the first place in the cells, set (8) consists of the elements which follow 2 in the first place in the cells; and group IV is assembled in the same way as group III except the numbers in the second place in the cells are used to select the elements for each set. Thus we have the configuration:



Set	Group I (rows)	Group II (columns)	Group III (first place)	Group IV (second place)
(1)	1 4 7	(4) 1 2 3	(7) 1 6 8	(10) 1 5 9
(2)	2 5 8	(5) 4 5 6	(8) 2 4 9	(11) 2 6 7
(3)	3 6 9	(6) 7 8 9	(9) 3 5 7	(12) 3 4 8

In the 12 sets of 3 elements, each of the 9 elements occurs with every other element once and only once in a set.

This is an illustration of one series of configurations which can be constructed with the aid of the completely orthogonalized squares. These are the  $EG(\kappa, p^n)$  in two dimensional space when  $\kappa = 2$  and  $p^n = 2, 3, 2^2, 5, 7, 2^3, 3^2, 11, 13, \dots$ . The  $PG(\kappa, p^n)$  configurations can be written by adding  $(k+1)$  elements to the previous group of configurations. For example, the elements 10, 11, 12 and 13 may be added to the groups, one to each group. That is, 10 is added to each set in group I, 11 is added to each set in group II, 12 to group III and 13 to group IV. An additional set must be added to include these four new elements. A configuration for  $t = 13, b = 13, k = 4, r = 4$  and  $\lambda = 1$  results.

Set

(1)	1 4 7 10	(4)	1 2 3 11	(7)	1 6 8 12	(10)	1 5 9 13
(2)	2 5 8 10	(5)	4 5 6 11	(8)	2 4 9 12	(11)	2 6 7 13
(3)	3 6 9 10	(6)	7 8 9 11	(9)	3 5 7 12	(12)	3 4 8 13
						(13)	10 11 12 13.

The 13 sets are made up of 4 elements each. These designs are symmetrical for sets and elements, that is, every pair of elements occurs together in the same number of sets, also, every pair of sets has the same number of elements in common. Discussion of the construction of these designs with illustrations are given in references [20, 8, 9] and [19].

In the  $PG(\kappa, p^n)$  series of designs, as constructed by means of completely orthogonalized squares, the sets cannot be arranged in replication groups. However, these configurations can be arranged in Youden squares [22] in which all the sets are placed side by side and all the elements in a single row form a complete replication. This method of arrangement has been of considerable value in experimentation with plants. The Youden squares are the  $PG(\kappa, p^n)$  when  $\kappa = 2$ . Singer [17] gives a partial list of the (reduced) perfect difference sets (table IV), only a single set for each  $p^n$ . The number of distinct perfect difference sets (or the number of distinct perfect partitions) for a given  $p^n$  is equal to  $\varphi(q)/3n$ . Since each perfect difference set can be paired with its inverse, the number is even.

The construction of one of the Youden squares from its perfect difference set will be illustrated. Consider  $p^n = 3$  then  $q = p^{2n} + p^n + 1 = 3^2 + 3 + 1 = 13$ . There are two perfect difference sets with their inverses for  $q = 13$ . One perfect difference set is 0, 1, 3, 9 which has the perfect partition 1, 2, 6, 4 which will add in succession to each number from 1 to and including 13, and also 1, 2, 6, 4

add to 13. The elements of the perfect difference set are put in set (1) except that 13 replaces 0. Set (2) is secured by a one-step cyclic substitution, 1 for 13, 2 for 1, 4 for 3 and 10 for 9. This process is continued until there are thirteen sets. If the substitution is applied to set (13), the elements in set (1) are secured.

		Set												
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Replica- tion	A	13	1	2	3	4	5	6	7	8	9	10	11	12
	B	1	2	3	4	5	6	7	8	9	10	11	12	13
	C	3	4	5	6	7	8	9	10	11	12	13	1	2
	D	9	10	11	12	13	1	2	3	4	5	6	7	8

This is the Youden square for  $t = 13$ ,  $b = 13$ ,  $r = 4$ ,  $k = 4$ , and  $\lambda = 1$ . The elements in each row form a complete replication.

TABLE IV  
*Singer's list of perfect difference sets*

$p^n$	$q$	$\varphi(q)$ $3^n$	Perfect difference set												
2	7	2 0 1 3													
2 <sup>2</sup>	21	2 0 1 4	14	16											
2 <sup>3</sup>	73	8 0 1 3	7	15	31	36	54	63							
2 <sup>4</sup>	273	12 0 1 3	7	15	31	63	90	116	127	136	181	194	204	233	238 255
3	13	4 0 1 3	9												
3 <sup>2</sup>	91	12 0 1 3	9	27	49	56	61	77	81						
5	31	10 0 1 3	8	12	18										
7	57	12 0 1 3	13	32	36	43	52								
11	133	36 0 1 3	12	20	34	38	81	88	94	104	109				
13	183	40 0 1 3	16	23	28	42	76	82	86	119	137	154	175		

$$t = q = p^{2n} + p^n + 1$$

A third series of configurations, called Lattice squares or quasi-Latin squares [21] can be constructed by using the completely orthogonalized squares. The groups of sets on page 78 are taken in pairs. For each pair a square is constructed having its rows formed by the sets of one group and its columns by the sets of another group. For example, square I below is made so that the sets of group I form the rows and the sets of group II form the columns. Square II is the combination of groups III and IV.

Square I

1	4	7
2	5	8
3	6	9

Square II

1	6	8
9	2	4
5	7	3

In this lattice square each pair of elements occurs together once only in either a row or a column of either one of the squares. Also, every element occurs with every other element once in one column and one row from each square.

A device known as "complements" gives several configurations. From an arrangement having  $k \neq \frac{1}{2}t$ , a second one can be obtained for the same number of elements, in sets of  $t - k$  units. This is done by replacing each set by its complement, that is, by a set containing all the elements missing from the original set. An illustration follows:

$$\begin{aligned} t &= 7, & b &= 7 \\ r &= 3, & k &= 3 \\ \lambda &= 1 \end{aligned}$$

$$\begin{aligned} t &= 7, & b &= 7 \\ r &= 4, & k &= 4 \\ \lambda &= 2 \end{aligned}$$

Set				Set			
(1)	1	2	4	(1)	3	5	6
(2)	2	3	5	(2)	1	4	6
(3)	3	4	6	(3)	1	2	5
(4)	4	5	7	(4)	1	2	3
(5)	5	6	1	(5)	2	3	4
(6)	6	7	2	(6)	1	3	4
(7)	7	1	3,	(7)	2	4	5

While the triple systems, quadruple systems, etc., which have been considered by some mathematicians, do furnish designs meeting the balance requirements, they are usually not suitable for experimental purposes. A quadruple system requires that every possible triple of elements occur once and only once together in a block. Since we need only every pair together once ( $\lambda = 1$ ) or more, only the triple systems are generally useful.

**4. Summary.** The mathematical theory of configuration has been helpful in the construction of the balanced incomplete block designs. It would be useful to know (a) what configurations (within the useful range) exist, (b) how these configurations may be constructed. In table I the configurations have been classified according to the value of  $\lambda$ , while in table II configurations within a useful range have been listed. Of the designs in this table which have not been constructed, some are known to exist. Those aids which have been used in the construction of the balanced incomplete block designs have been briefly discussed.

#### REFERENCES

- [1] BALL, W. W. R. Revised by COXETER, H. S. M. *Mathematical recreations and essays*, Macmillan and Co., London, 11th Edition, 1939.
- [2] BOSE, R. C. "On the Application of the Properties of Galois Fields to the Problem of the Construction of Hyper-Graeco Latin Squares," *Sankhyā*, Vol. 3(1938), pp. 323-338.
- [3] BOSE, R. C. AND NAIR, K. R. "Partially Balanced Incomplete Block Designs," *Sankhyā*, Vol. 4(1939), pp. 337-372.

- [4] BOSE, R. C. "On the Construction of Balanced Incomplete Block Designs," *Ann. of Eugenics*, Vol. 9 Part 4 (in the press).
- [5] CARMICHAEL, R. D. *Introduction to the theory of groups of finite order*, Ginn and Company, 1937.
- [6] COLE, F. N. "Kirkman Parades," *Bull. of the Am. Math. Society*, Vol. 28(1922), pp. 435-437.
- [7] ECKENSTEIN, O. "Bibliography of Kirkman's School-girl Problem," *Messenger of Mathematics*, Vol. 41(1911), pp. 33-36.
- [8] FISHER, R. A. AND YATES, F. *Statistical tables for biological, agricultural, and medical research*, Oliver and Boyd, Edinburgh, 1938.
- [9] GOULDEN, C. H. "Modern Methods for Testing a Large Number of Varieties," *Dominion of Canada, Department of Agriculture, Tech. Bull.* 9, 1937.
- [10] GOULDEN, C. H. *Methods of statistical analysis*, John Wiley and Sons, New York, 1939.
- [11] KIRKMAN, T. P. "On a Problem in Combinations," *Cambridge and Dublin Math. Jour.*, Vol. 2(1847), pp. 191-204.
- [12] KIRKMAN, T. P. "Query," *Lady's and Gentleman's diary*, p. 48, 1850.
- [13] MOORE, E. H. "Concerning Triple Systems," *Math. Ann.*, Vol. 43(1893), pp. 271-285.
- [14] MULDER, P. *Kirkman-Systemen* (Groningen Dissertation), Leiden, 1917.
- [15] NETTO, EUGEN. *Lehrbuch der combinatorik*. Verlag und Druck von B. G. Teubner, Leipzig und Berlin, 1927. Zweite Auflage.
- [16] REISS, M. "Ueber eine Steinersche combinatorische Aufgabe, welche im 45<sup>sten</sup> Bande dieses Journals, Seite 181, gestellt worden ist," *Crelles Journal*, Vol. 56(1859), pp. 326-344.
- [17] SINGER, JAMES. "A Theorem in Finite Projective Geometry and some Applications to Number Theory," *Trans. Am. Math. Society*, Vol. 43(1938), pp. 377-385.
- [18] STEVENS, W. L. "The Completely Orthogonalized Latin Square," *Ann. of Eugenics*, Vol. 9(1939), pp. 82-93.
- [19] WEISS, MARTIN G. AND COX, GERTRUDE M. "Balanced Incomplete Block and Lattice Square Designs for Testing Yield Differences Among Large Numbers of Soybean Varieties," *Iowa Agr. Exp. Sta. Res. Bull.*, 257(1939), pp. 293-316.
- [20] YATES, F. "Incomplete Randomized Blocks," *Ann. of Eugenics*, Vol. 7(1936), pp. 121-140.
- [21] YATES, F. "A Further Note on the Arrangement of Variety Trials: quasi-Latin Squares," *Ann. of Eugenics*, Vol. 7(1937), pp. 319-332.
- [22] YOUNDEN, W. J. "Use of Incomplete Block Replications in Estimating Tobacco-mosaic Virus," *Contributions from Boyce Thompson Inst.*, Vol. 9(1937), pp. 41-48.

IOWA STATE COLLEGE,  
AMES, IOWA.

# A COMPARISON OF ALTERNATIVE TESTS OF SIGNIFICANCE FOR THE PROBLEM OF $m$ RANKINGS<sup>1</sup>

BY MILTON FRIEDMAN

A paper published in 1937 [2] suggested that the consilience of a number of sets of ranks can be tested by computing a statistic designated  $\chi_r^2$ . A mathematical proof by S. S. Wilks demonstrated that the distribution of  $\chi_r^2$  approaches the ordinary  $\chi^2$  distribution as the number of sets of ranks increases. The rapidity with which this limiting distribution is approached was investigated by obtaining the exact distributions of  $\chi_r^2$  for a number of special cases. It was concluded that "when the number of sets of ranks is moderately large (say greater than 5 for four or more ranks) the significance of  $\chi_r^2$  can be tested by reference to the available  $\chi^2$  tables" [2, p. 695]. The use of the normal distribution was recommended when the number of ranks in each set is large, but the number of sets of ranks is small, although no rigorous justification of this procedure was presented.

Except for the few special cases for which exact distributions were given, the paper did not provide a test of significance for data involving less than six sets of ranks and a small or moderate number of ranks in each set. This important gap has now been filled by M. G. Kendall and B. Babington Smith [1]. In addition, they furnish a somewhat more exact test of significance for tables of ranks for which the earlier article recommended the use of the  $\chi^2$  distribution.

Kendall and Smith use a different statistic,  $W$ , defined as  $\chi_r^2$  divided by its maximum value,  $m(n-1)$ , where  $n$  is the number of items ranked, and  $m$  the number of sets of ranks.<sup>2</sup> The new statistic (independently suggested by W. Allen Wallis [3] who terms it the rank correlation ratio and denotes it by  $\eta_r^2$ ) is thus not fundamentally different from  $\chi_r^2$ . A more radical innovation is the improvement in the test of significance that they suggest. Instead of testing  $\chi_r^2$  by reference to the  $\chi^2$  distribution for  $n-1$  degrees of freedom, Kendall and Smith, generalizing from the first four moments of  $W$ , recommend that the significance of  $W$  be tested by reference to the analysis of variance distribution (Fisher's  $z$ -distribution) with  $z = \frac{1}{2} \log_e \left( \frac{(m-1)W}{1-W} \right)$ ,  $n_1 = (n-1) - \frac{2}{m}$ ,  $n_2 = (m-1) \left[ (n-1) - \frac{2}{m} \right]$ . For small values of  $m$  and  $n$ , they introduce con-

<sup>1</sup> The author is indebted to Mr. W. Allen Wallis for valuable criticism and to Miss Edna R. Ehrenberg for computational assistance.

<sup>2</sup> This is Kendall and Smith's notation which will be used in the present paper. The original paper [2] designated the number of items ranked by  $p$ , and the number of sets of ranks by  $n$ .

tinuity corrections, substituting for  $W = \frac{12S}{m^2(n^3 - n)}$ , the statistic

$$W_c = \frac{\frac{S - 1}{m^2(n^3 - n)} + 2}{12} = \frac{W - \frac{12}{m^2(n^3 - n)}}{1 + \frac{24}{m^2(n^3 - n)}},$$

where  $S$  is the observed sum of squares of the deviations of sums of ranks from the mean value,  $m(n + 1)/2$ . Comparison with exact distributions of  $W$  (or  $S$ ) for special cases indicates that this test yields very good approximations to the correct probabilities.

In the limit the two tests of significance are identical. Neglecting the correction for continuity,  $z = \frac{1}{2} \log_e \left( \frac{(m - 1)\chi_r^2}{m(n - 1) - \chi_r^2} \right) \rightarrow \frac{1}{2} \log_e \left( \frac{\chi_r^2}{n - 1} \right)$ ,  $n_2 = (m - 1) \left[ (n - 1) - \frac{2}{m} \right] \rightarrow \infty$ , and  $n_1 = (n - 1) - \frac{2}{m} \rightarrow (n - 1)$  as  $m \rightarrow \infty$ . For  $n_2 = \infty$ , the analysis of variance distribution is identical with the distribution of  $\frac{1}{2} \log_e \frac{\chi^2}{n_1}$ . The difference between the two tests is thus that one,  $\chi^2$ , uses a single (limiting) distribution for all values of  $m$ , whereas the other,  $z$ , adapts the distribution to the value of  $m$ .

The necessity of taking into account the value of  $m$ , while it increases the flexibility of the distribution, makes the  $z$  test somewhat less convenient in practice than the  $\chi^2$  test. Additional computation is required to obtain the values of  $n_1$  and  $n_2$ , and to make the continuity corrections. It is also fairly laborious to test the significance of the result, if exact values of  $z$  at any level of significance are required. In these instances, two-way interpolation of reciprocals in the analysis of variance tables is necessary since both  $n_1$  and  $n_2$  are always fractional. These difficulties make it desirable to investigate the rapidity with which the significance levels given by the  $z$  test approach those given by the  $\chi^2$  test, and thus determine the range of values of  $m$  and  $n$  for which the simpler test can safely be employed. This investigation will yield as a by product the .05 and .01 significance values of  $\chi_r^2$  (or  $W$  or  $S$ ) for selected values of  $m$  and  $n$  as determined by the  $z$  test.

Table I presents a summary comparison of the values of  $\chi_r^2$  at the .05 and .01 levels of significance as shown by (1) exact distributions, (2) the  $z$  test with continuity corrections, (3) the  $\chi^2$  test.<sup>3</sup> The significance values are expressed in terms of  $\chi_r^2$  rather than  $W$  because, for a given number of ranks per set (i.e., a given  $n$ ), the significance values given by the  $\chi^2$  test are the same regardless of the number of sets of ranks (i.e., of the value of  $m$ ). This would not be so if  $W$  were employed, since  $W = \chi_r^2/m(n - 1)$ . The expected value of  $W$  depends on

<sup>3</sup> The values of  $\chi_r^2$  computed using the  $z$  test that are given in Tables I and II were obtained with the aid of Fisher and Yates' Table V [4]. Linear interpolation of reciprocals was employed throughout.

$m$  and approaches zero as  $m \rightarrow \infty$  while the expected value of  $\chi^2$  is equal to  $n - 1$  for all values of  $m$ .

The values given by the  $z$  test agree remarkably well with the exact values. With but two exceptions (the .01 values for  $n = 3$ ,  $m = 8$  and 10) the exact value differs very much less from the value given by the  $z$  test than from the value given by the  $\chi^2$  test. In all but three of the 12 comparisons, the  $z$  test gives a value below the correct one.<sup>4</sup>

TABLE I

*Comparison of Values of  $\chi^2_r$  at .05 and .01 Levels of Significance Yielded by Exact Distributions,  $z$  Test with Continuity Corrections, and  $\chi^2$  Test*

$n$	$m$	.05 Level of Significance				.01 Level of Significance			
		From Exact Distribution		From $z$ test with continuity corrections	From $\chi^2$ test	From Exact Distribution		From $z$ test with continuity corrections	From $\chi^2$ test
		Limits	Interpolated value*			Limits	Interpolated value*		
3	8	5.25-6.25	6.16	6.012	5.991		9.00	8.35	9.21
	9	6.0 -6.22	6.17	6.004	5.991		8.67	8.44	9.21
	10	5.6 -6.2	6.08	5.999	5.991	8.6 - 9.6	9.04	8.51	9.21
	$\infty$			5.991	5.991			9.21	9.21
4	4	7.5 -7.8	7.54	7.43	7.82	9.3 - 9.6	9.42	9.21	11.34
	5	7.32-7.8	7.54	7.52	7.82	9.72- 9.96	9.87	9.66	11.34
	6	7.4 -7.6	7.49	7.57	7.82		10.00	9.95	11.34
	$\infty$			7.82	7.82			11.34	11.34
5	3	8.27-8.53	8.41	8.59	9.49	9.87-10.13	10.05	10.08	13.28
	$\infty$			9.49	9.49			13.28	13.28

\* Computed by linear interpolation of probabilities.

Table II gives for a very much larger number of values of  $m$  and  $n$  the .05 and .01 values of  $\chi^2_r$  computed on the basis of the  $z$  test with continuity correc-

<sup>4</sup> These comparisons duplicate some of those made by Kendall and Smith and merely serve to confirm their conclusion that the  $z$  test with continuity corrections gives exceedingly good results.

The values obtained using the  $z$  test without continuity corrections agree less well with the exact values than those obtained with the aid of the continuity corrections. However even if no continuity corrections are made the  $z$  test in general yields values closer to the exact values than does the  $\chi^2$  test.

TABLE II

*Values of  $\chi^2$  at .05 and .01 Levels of Significance Computed on the Basis of Kendall and Smith's  $z$  test, with Continuity Corrections; .10, .075, .02, .015 Values of  $\chi^2$*

$m$	$n$				
	3	4	5	6	7
Values at .05 Level of Significance					
3			8.59	9.90	11.24
4		7.43	8.84	10.24	11.62
5		7.52	8.98	10.42	11.84
6		7.57	9.08	10.54	11.97
8	6.012	7.63	9.18	10.68	12.14
10	5.999	7.67	9.25	10.76	12.23
15	5.985	7.72	9.33	10.87	12.36
20	5.983	7.74	9.37	10.92	12.42
100	5.987	7.80	9.46	11.04	12.56
$\infty$	5.991	7.82	9.49	11.07	12.59
$\chi^2 (.10)$	4.605	6.25	7.78	9.24	10.64
$\chi^2 (.075)^*$	5.18	6.90	8.49	10.00	11.45
Values at .01 Level of Significance					
3			10.08	11.69	13.26
4		9.21	10.93	12.59	14.19
5		9.66	11.42	13.11	14.74
6		9.95	11.74	13.45	15.09
8	8.35	10.31	12.13	13.87	15.53
10	8.51	10.52	12.37	14.11	15.79
15	8.74	10.79	12.67	14.44	16.14
20	8.85	10.93	12.82	14.60	16.31
100	9.14	11.26	13.19	14.99	16.71
$\infty$	9.21	11.34	13.28	15.09	16.81
$\chi^2 (.02)$	7.82	9.84	11.67	13.39	15.03
$\chi^2 (.015)^*$	8.40	10.46	12.34	14.09	15.77

\* Computed from Fisher and Yates' Table IV (4) by linear interpolation between the logarithms of the probabilities.



tions. The values entered for  $m = \infty$  are obtained from  $\chi^2$  tables for  $n - 1$  degrees of freedom and are the significance values by the  $\chi^2$  test for all values of  $m$ . It is apparent that as  $m$  increases the .01 and .05 values of  $\chi_r^2$  approach their limiting values very rapidly. For  $n = 7$ , two-thirds of the difference between the .05 values for  $m = 3$  and  $m = \infty$ , and an even larger proportion of the difference between the .01 values, disappears by the time  $m = 10$ ; and the situation is similar for the other values of  $n$ . Except for the .05 values for  $n = 3$ , the approach to the limit is monotonic from below. The use of the  $\chi^2$  test thus tends to lead to the overestimation of the significance values and of the probabilities attached to observed values of  $\chi_r^2$ . It is clear, however, that for large and even moderate values of  $m$  the  $\chi^2$  test is, for all practical purposes, equivalent to the  $z$  test.

In order to determine more precisely the range of values of  $m$  and  $n$  for which the approximation given by the  $\chi^2$  test is adequate, it is necessary to adopt some convention about the error in estimated significance values of  $\chi_r^2$  that is tolerable. Since the conclusion drawn from an observed  $\chi_r^2$  depends on the probability that it will be exceeded by chance, this convention clearly should be expressed in terms of the error in the probability.

The structure of published  $\chi^2$  tables makes it convenient to accept an estimated probability between .10 and .05 as a tolerable approximation to a correct probability of .05, and an estimated probability between .02 and .01 as a tolerable approximation to a correct probability of .01. These ranges of tolerance are entirely on one side of the correct probability because, as pointed out above, the error in using the  $\chi^2$  test is consistent in direction. These ranges are purely arbitrary, of course, and many may think them too broad.

On the basis of this or some similar convention it is possible to make objective statements concerning the range of values of  $m$  and  $n$  for which the  $\chi^2$  test is adequate. The next to the last line in the first section of Table II gives the .10 values of  $\chi^2$ ; the next to the last line in the second section, the .02 values. All the .05 values of  $\chi_r^2$  shown in the table exceed the .10 value of  $\chi^2$ . Using the  $\chi^2$  test, all of the values (with two exceptions for  $n = 3$ ) would signify a probability greater than .05 but less than .10. Thus the error made at the .05 level is within the admissible range according to the suggested convention. The  $\chi^2$  test is therefore an adequate substitute for the  $z$  test at the .05 level for all values of  $m$  and  $n$  except possibly for a few of the values for which exact distributions are available.

As might be expected, the  $\chi^2$  test is less satisfactory at the .01 level. For values of  $m$  less than six, the .01 values of  $\chi_r^2$  computed using the  $z$  test with continuity corrections are less than the .02 value of  $\chi^2$ . For  $m$  greater than 5, the values of  $\chi_r^2$  in the table would all be accorded a probability greater than .01 but less than .02 if the  $\chi^2$  test were employed. As already noted, this is the range of values of  $m$  for which the original paper suggested the  $\chi^2$  test could validly be used [2, p. 695].

In view of the arbitrary nature of the convention as to the permissible error

in the probability attached to an observed value of  $\chi_r^2$ , it is interesting to investigate the effect of an alternative and stricter convention, namely, that only probabilities from .075 to .05 and from .015 to .01 be accepted as approximations to correct probabilities of .05 and .01 respectively. The .075 and .015 values of  $\chi^2$  are given in the last lines of the two sections of Table II. On the basis of this convention the  $\chi^2$  test is adequate at the .05 level for  $m$  greater than three, and

TABLE III

*Values of  $S$  at .05 and .01 Levels of Significance Computed on the Basis of Kendall and Smith's  $z$  test, with Continuity Corrections*

$m$	$n$					Additional values for $n = 3$	
	3	4	5	6	7	$m$	$S$
Values at .05 Level of Significance							
3			64.4	103.9	157.3	9	54.0
4		49.5	88.4	143.3	217.0	12	71.9
5		62.6	112.3	182.4	276.2	14	83.8
6		75.7	136.1	221.4	335.2	16	95.8
8	48.1	101.7	183.7	299.0	453.1	18	107.7
10	60.0	127.8	231.2	376.7	571.0		
15	89.8	192.9	349.8	570.5	864.9		
20	119.7	258.0	468.5	764.4	1158.7		
Values at .01 Level of Significance							
3			75.6	122.8	185.6	9	75.9
4		61.4	109.3	176.2	265.0	12	103.5
5		80.5	142.8	229.4	343.8	14	121.9
6		99.5	176.1	282.4	422.6	16	140.2
8	66.8	137.4	242.7	388.3	579.9	18	158.6
10	85.1	175.3	309.1	494.0	737.0		
15	131.0	269.8	475.2	758.2	1129.5		
20	177.0	364.2	641.2	1022.2	1521.9		

at the .01 level for  $m$  greater than nine, except possibly for a few of the values for which exact distributions are available. Thus even so drastic a lowering of the permissible margin of error as halving it limits only slightly the range of values of  $m$  for which the  $\chi^2$  test is adequate.

Table II provides, of course, a direct means of testing the significance of observed values of  $\chi_r^2$  for the tabled values of  $m$  and  $n$ . For this purpose, however, Table III, giving the significance values of  $S$  is more useful, since it obviates

the necessity of converting  $S$  into  $\chi^2$ . For  $n = 3$  Table III includes a few values of  $m$  in addition to those in Table II.

#### SUMMARY

The preceding analysis suggests that the  $\chi^2$  test of the significance of  $\chi^2$  (or  $W$  or  $\eta^2$ ), while less accurate than the  $z$  test proposed by Kendall and Smith, is adequate for practical purposes at the .01 level of significance if the number of sets of ranks ( $m$ ) is greater than 5; and at the .05 level for any number of sets of ranks, provided the number of ranks in each set ( $n$ ) is more than 3. Exact distributions are now available for  $n = 3$ ,  $m = 3$  to 10;  $n = 4$ ,  $m = 3$  to 6;  $n = 5$ ,  $m = 3$  [1]. The .05 and .01 values of  $\chi^2$  and  $S$ , computed using the Kendall and Smith  $z$  test with continuity corrections, are given in Tables II and III of the present note for  $n = 3$  to 7 and selected values of  $m$  from 3 to 100. For  $n$  greater than 7 and  $m$  less than 6, the  $z$  test with continuity corrections should be employed. For all other combinations of  $n$  and  $m$  not covered by the exact distributions or by Tables II and III, the  $\chi^2$  test is adequate.

#### REFERENCES

- [1] M. G. Kendall and B. Babington Smith, "The Problem of  $m$  Rankings," *Annals of Math. Stat.*, Vol. X (1939), pp. 275-87.
- [2] Milton Friedman, "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," *Jour. Am. Stat. Assn.*, Vol. XXXII (1937), pp. 675-701.
- [3] W. Allen Wallis, "The Correlation Ratio for Ranked Data," *Jour. Am. Stat. Assn.*, Vol. XXXIV (1939), pp. 533-8.
- [4] R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (London: Oliver and Boyd, 1938).

NATIONAL BUREAU OF ECONOMIC RESEARCH,  
NEW YORK.

## NOTES

*This section is devoted to brief research and expository articles, notes on methodology and other short items.*

### NOTE ON AN APPROXIMATE FORMULA FOR THE SIGNIFICANCE LEVELS OF $Z$

BY W. G. COCHRAN

1. **Introduction.** An important part has been played in modern statistical analysis by the distribution of  $z = \frac{1}{2} \log \frac{s_1^2}{s_2^2}$ , when  $s_1^2$  and  $s_2^2$  are two independent estimates of the same variance. In particular, all tests of significance in the analysis of variance and in multiple regression problems are based on this distribution. Complete tabulation of the frequency distribution of  $z$  is a heavy task, because the distribution is a two-parameter one, the parameters being the number of degrees of freedom,  $n_1$  and  $n_2$  in the estimates  $s_1^2$  and  $s_2^2$ . Thus each significance level of  $z$  requires a separate two-way table. Fisher constructed a table of the 5 percent points in 1925 [1], and this has since been extended by several workers [2] to the 20, 1, and 0.1 percent level for a somewhat wider range of values of  $n_1$  and  $n_2$ .

With his original table, Fisher gave an approximate formula for the 5 percent values of  $z$ , for high values of  $n_1$  and  $n_2$  outside the limits of his table. The formula reads:

$$(1) \quad z \text{ (5 percent)} = \frac{1.6449}{\sqrt{h-1}} - 0.7843 \left( \frac{1}{n_1} - \frac{1}{n_2} \right),$$

$$\text{where } \frac{2}{h} = \frac{1}{n_1} + \frac{1}{n_2}.$$

The constant 1.6449 is the 5 percent significance level for a *single tail* of the normal distribution, and the constant 0.7843 will be found to be  $\frac{1}{6}\{2 + (1.6449)^2\}$ . Thus the general formula for the significance levels of  $z$  derivable from (1) is

$$= \frac{x}{\sqrt{h-1}} - \left( \frac{x^2 + 2}{6} \right) \left( \frac{1}{n_1} - \frac{1}{n_2} \right),$$

where  $x$  is a normal deviate with unit standard error. By inserting the appropriate significance level of  $x$ , this formula has been extended [2] to the tables of the 20, 1, and 0.1 percent levels of  $z$  and commonly appears with all published tables of  $z$ . The objects of this note are to indicate the derivation of the formula and to suggest an improvement upon it in the latter cases.

**2. The transformation of the  $z$ -distribution to normality.** For high values of  $n_1$  and  $n_2$ , the distribution of  $z$  approaches the normal distribution, the principal deviation being a slight skewness introduced by the inequality of  $n_1$  and  $n_2$ . It is therefore natural to seek an approximate formula for the distribution of  $z$  by examining its relation to the normal distribution. For the  $z$ -distribution the ratio  $\kappa_r/\kappa_2^{r/2}$ , where  $\kappa_r$  is the  $r^{\text{th}}$  cumulant, is of the order  $n^{-(r-1)}$ , where  $n$  is the smaller of  $n_1$  and  $n_2$ . This property is common to a large number of distributions which tend to normality; for example, the distribution of the mean of a sample of size  $n$  from any distribution with finite cumulants. Fisher and Cornish [3] have recently given a method, applicable to all distributions with this property, for transforming the distribution to a normal distribution to any desired order of approximation. They also obtained explicit expressions for the significance levels of the original distribution in terms of the significance levels of the normal distribution, discussing the  $z$ -distribution as a particular example. The relation between  $z$  and the normal deviate  $x$  at the same level of probability was found to be

$$(2) \quad z = \frac{x}{\sqrt{h}} - \frac{1}{6}(x^2 + 2)\left(\frac{1}{n_1} - \frac{1}{n_2}\right) + \frac{1}{\sqrt{h}}\left\{\frac{x^3 + 3x}{12h} + \frac{x^3 + 11x}{144}h\left(\frac{1}{n_1} - \frac{1}{n_2}\right)^2\right\},$$

the three terms on the right hand side being respectively of order  $n^{-1}$ ,  $n^{-1}$ , and  $n^{-1}$ , so that terms of order  $n^{-2}$  are neglected.<sup>1</sup>

If this equation is compared with equation (1), the latter appears at first sight to be the approximation of order  $n^{-1}$  to the  $z$ -distribution, except that the divisor of  $x$  is  $\sqrt{h-1}$  in (1) and  $\sqrt{h}$  in (2). Computation of a few values shows that at the 5 percent level, equation (1) is the better approximation. For example, for  $n_1 = 40$ ,  $n_2 = 60$ , (1) gives  $z$  (5 percent) = .2334, (2) gives .2309, and the exact value is .2332.

Since

$$\frac{x}{\sqrt{h-1}} = \frac{x}{\sqrt{h}} + \frac{x}{2h\sqrt{h}} + \text{terms of order } n^{-2},$$

Fisher's approximation differs from (2) by including a correction term of order  $n^{-1}$ . Inspection of the true correction terms of this order in equation (2) shows that for finite values of  $n_1$  and  $n_2$  the term  $\frac{x^3 + 11x}{144} \sqrt{h} \left(\frac{1}{n_1} - \frac{1}{n_2}\right)^2$  is considerably smaller than the term  $\frac{x^3 + 3x}{12h\sqrt{h}}$ , since the former has a smaller numerical

coefficient and involves the difference between  $\frac{1}{n_1}$  and  $\frac{1}{n_2}$ . Thus Fisher's formula gives a close approximation to the true formula of order  $n^{-1}$ , provided that  $\frac{x}{2}$  is approximately equal to  $\frac{x^3 + 3x}{12}$ ; i.e. if  $\frac{x^2 + 3}{6}$  is approximately equal

<sup>1</sup> Fisher and Cornish also gave the two succeeding terms.

to 1. For the 5 percent level,  $x = 1.6449$ , and  $\frac{x^2 + 3}{6} = 0.951$ . Thus at the 5 percent level the use of  $\sqrt{h-1}$  in (1) instead of  $\sqrt{h}$  extends the validity of Fisher's approximation from order  $n^{-1}$  to order  $n^{-1}$ .

This ingenious device, however, requires adjustment at other levels of significance. The values of  $(x^2 + 3)/6$  at the principal significance levels are shown below.

Significance level—%	40	30	20	10	5	1	0.1
$\lambda = (x^2 + 3)/6$	0.51	0.55	0.62	0.77	0.95	1.40	2.09

If  $\sqrt{h-1}$  in formula (1) is replaced by  $\sqrt{h-\lambda}$ , with the above values of  $\lambda$ , Fisher's formula will be approximately valid to order  $n^{-1}$  at all levels of significance. In particular, for the tables already published of the 20, 1 and 0.1 percent points,  $\lambda$  may be taken as 0.6, 1.4 and 2.1 respectively. The values of  $z$  given by the use of  $\sqrt{h-1}$  and  $\sqrt{h-\lambda}$  are compared below for  $n_1 = 24$ ,  $n_2 = 60$ .<sup>2</sup>

Significance Level	Approximate formula		Exact value
	$\sqrt{h-1}$	$\sqrt{h-\lambda}$	
20%	.1346	.1337	.1338
1%	.3723	.3748	.3746
0.1%	.4875	.4966	.4955

The use of  $\sqrt{h-\lambda}$  gives values practically correct to 4 decimal places, except for the 0.1 level of significance, at which the higher terms become more important.

With the aid of this formula, complete tabulation of the  $z$ -distribution for a given pair of high values of  $n_1$  and  $n_2$  is relatively simple. If very low probabilities at the tails are required, the further approximations given by Fisher and Cornish [3] may be used.

#### REFERENCES

- [1] R. A. FISHER. *Statistical Methods for Research Workers*. Edinburgh, Oliver and Boyd. 1st Ed. 1925.
- [2] R. A. FISHER AND F. YATES. *Statistical Tables*. Edinburgh, Oliver and Boyd. 1938.
- [3] E. A. CORNISH AND R. A. FISHER. "Moments and Cumulants in the Specification of Distributions," *Revue de l'Institut International de Statistique*, Vol. 4 (1937).

IOWA STATE COLLEGE,  
AMES, IOWA.

<sup>2</sup> The numerical terms in the approximate formula given for the 20 percent points on p. 28 of Fisher and Yates' *Statistical Tables* are in error. Their formula should read:

$$z = \frac{0.8416}{\sqrt{h-1}} - 0.4514 \left( \frac{1}{n_1} - \frac{1}{n_2} \right)$$

# A NOTE ON THE ANALYSIS OF VARIANCE WITH UNEQUAL CLASS FREQUENCIES<sup>1</sup>

BY ABRAHAM WALD<sup>2</sup>

Let us consider  $p$  groups of variates and denote by  $m_j$  ( $j = 1, \dots, p$ ) the number of elements in the  $j$ -th group. Let  $x_{ij}$  be the  $i$ -th element of the  $j$ -th group. We assume that  $x_{ij}$  is the sum of two variates  $\epsilon_{ij}$  and  $\eta_j$ , i.e.  $x_{ij} = \epsilon_{ij} + \eta_j$ , where  $\epsilon_{ij}$  ( $i = 1, \dots, m_j$ ;  $j = 1, \dots, p$ ) is normally distributed with mean  $\mu$  and variance  $\sigma^2$ , and  $\eta_j$  ( $j = 1, \dots, p$ ) is normally distributed with mean  $\mu'$  and variance  $\sigma'^2$ . All the variates  $\epsilon_{ij}$  and  $\eta_j$  are supposed to be distributed independently.

The intraclass correlation  $\rho$  is given by<sup>3</sup>

$$\rho = \frac{\sigma'^2}{\sigma^2 + \sigma'^2}.$$

Confidence limits for  $\rho$  have been derived only in case of equal class frequencies, i.e.  $m_1 = m_2 = \dots = m_p$ . In this paper we shall deal with the problem of determining the confidence limits for  $\rho$  in the case of unequal class frequencies.

Since  $\rho$  is a monotonic function of  $\frac{\sigma'^2}{\sigma^2}$ , our problem is solved if we derive confidence limits for  $\frac{\sigma'^2}{\sigma^2}$ .

Denote by  $\bar{x}_j$  the arithmetic mean of the  $j$ -th group, i.e.

$$(1) \quad \bar{x}_j = \frac{\sum_{i=1}^{m_j} \epsilon_{ij}}{m_j} + \eta_j.$$

Hence the variance of  $\bar{x}_j$  is equal to

$$(2) \quad \sigma_{\bar{x}_j}^2 = \frac{\sigma^2}{m_j} + \sigma'^2.$$

Denote  $\frac{\sigma'^2}{\sigma^2}$  by  $\lambda^2$ . Then we have

$$(3) \quad \sigma_{\bar{x}_j}^2 = \sigma^2 \left( \frac{1}{m_j} + \lambda^2 \right) \quad \frac{\sigma^2}{w_j}$$

<sup>1</sup>The author is indebted to Professor H. Hotelling for formulating the problem dealt with in this paper.

<sup>2</sup>Research under a grant-in-aid from the Carnegie Corporation at New York.

<sup>3</sup>See for instance R. A. Fisher, *Statistical Methods for Research Workers*, 6-th edition, p. 228.

where

$$(4) \quad w_j = \frac{m_j}{1 + m_j \lambda^2}.$$

Now we shall prove that

$$(5) \quad \frac{\sum_{j=1}^p w_j \bar{x}_j^2}{\sum_{j=1}^p w_j}$$

has the  $\chi^2$ -distribution with  $p - 1$  degrees of freedom. Let

$$y_j = \sqrt{w_j} \bar{x}_j \quad (j = 1, \dots, p)$$

and consider the orthogonal transformation

$$y'_1 = L_1(y_1, \dots, y_p),$$

$$y'_{p-1} = L_{p-1}(y_1, \dots, y_p),$$

$$y'_p = L_p(y_1, \dots, y_p) = \frac{\sqrt{w_1} y_1 + \dots + \sqrt{w_p} y_p}{\sqrt{w_1 + \dots + w_p}},$$

where  $L_1(y_1, \dots, y_p), \dots, L_{p-1}(y_1, \dots, y_p)$  denote arbitrary homogenous linear functions subject to the only condition that the transformation should be orthogonal.

Since the mean value of  $y_j$  is equal to  $\sqrt{w_j}(\mu + \mu')$  and the variance of  $y_j$  is equal to  $\sigma^2$ , we obviously have: The mean value of  $y'_j$  ( $j = 1, \dots, p - 1$ ) is equal to zero, the variance of  $y'_j$  ( $j = 1, \dots, p$ ) is equal to  $\sigma^2$ . In order to prove our statement, we have only to show that the expression (5) is equal to  $\frac{1}{\sigma^2}(y'^2_1 + \dots + y'^2_{p-1})$ . If we substitute in (5)  $\frac{y_j}{\sqrt{w_j}}$  for  $\bar{x}_j$ , we get

$$\begin{aligned} & \frac{1}{\sigma^2} \sum_{j=1}^p \left\{ w_j \frac{y_j^2}{w_j} - 2 \frac{y_j}{\sqrt{w_j}} \frac{\sum_{i=1}^p \sqrt{w_i} y_i}{\sum_{i=1}^p w_i} + \left( \frac{\sum_{i=1}^p \sqrt{w_i} y_i}{\sum_{i=1}^p w_i} \right)^2 \right\} \\ (5') \quad &= \frac{1}{\sigma^2} \sum_{j=1}^p y_j^2 - 2 \frac{(\sum_{i=1}^p \sqrt{w_i} y_i)^2}{\sum_{i=1}^p w_i} + \frac{(\sum_{i=1}^p \sqrt{w_i} y_i)^2}{\sum_{i=1}^p w_i} \\ & \quad \sum_{j=1}^p y_j^2 - \frac{(\sum_{i=1}^p \sqrt{w_i} y_i)^2}{\sum_{i=1}^p w_i} \quad \frac{1}{\sigma^2} \left[ \sum_{j=1}^p y_j^2 - y_p'^2 \right] = \frac{1}{\sigma^2} \left[ \sum_{j=1}^p y_j'^2 - y_p'^2 \right] \\ & \quad \frac{1}{\sigma^2} (y_1'^2 + \dots + y_{p-1}'^2). \end{aligned}$$



Since  $\frac{\Sigma \Sigma (x_{ij} - \bar{x}_i)^2}{\sigma^2}$  has the  $\chi^2$  distribution with  $N - p$  degrees of freedom, the expression

$$(6) \quad F = \frac{N - p}{p - 1} \frac{\sum_{i=1}^p \left\{ w_i \left( \bar{x}_i - \frac{\Sigma w_i \bar{x}_i}{\Sigma w_i} \right)^2 \right\}}{\Sigma \Sigma (x_{ij} - \bar{x}_i)^2}$$

has the analysis of variance distribution with  $p - 1$  and  $N - p$  degrees of freedom, where  $N = m_1 + \dots + m_p$ . In case  $m_1 = m_2 = \dots = m_p = m$ , we have

$$(6') \quad F = \frac{N - p}{p - 1} \frac{\sum_{i=1}^p (\bar{x}_i - \bar{x})^2}{\Sigma \Sigma (x_{ij} - \bar{x}_i)^2} \cdot \frac{m}{1 + m\lambda^2} = \frac{1}{1 + m\lambda^2} F^*,$$

where  $\bar{x} = \frac{\Sigma \Sigma x_{ij}}{N}$  and  $F^* = \frac{N - p}{p - 1} \frac{m \Sigma (\bar{x}_i - \bar{x})^2}{\Sigma \Sigma (x_{ij} - \bar{x}_i)^2}$ .

Hence

$$\lambda^2 = \left( \frac{F^*}{F} - 1 \right) \frac{1}{m}.$$

If  $F_1$  denotes the lower and  $F_2$  the upper confidence limit of  $F$ , we obtain for  $\lambda^2$  the confidence limits

$$\left( \frac{F^*}{F_1} - 1 \right) \frac{1}{m} \quad \text{and} \quad \left( \frac{F^*}{F_2} - 1 \right) \frac{1}{m}.$$

Let us now consider the general case that  $m_1, \dots, m_p$  are arbitrary positive integers. First we shall show that the set of values of  $\lambda^2$ , for which (6) lies between its confidence limits  $F_1$  and  $F_2$ , is an interval. For this purpose we have only to show that

$$f(\lambda^2) = \sum_{i=1}^p \left\{ w_i \left( \bar{x}_i - \frac{\Sigma w_i \bar{x}_i}{\Sigma w_i} \right)^2 \right\}$$

is monotonically decreasing with  $\lambda^2$ . In fact

$$\frac{df(\lambda^2)}{d\lambda^2} = \sum_{i=1}^p \frac{dw_i}{d\lambda^2} \left( \bar{x}_i - \frac{\Sigma w_i \bar{x}_i}{\Sigma w_i} \right)^2 - 2 \frac{d}{d\lambda^2} \left( \frac{\Sigma w_i \bar{x}_i}{\Sigma w_i} \right) \left[ \sum_{i=1}^p w_i \left( \bar{x}_i - \frac{\Sigma w_i \bar{x}_i}{\Sigma w_i} \right) \right].$$

Since

$$\sum_{i=1}^p w_i \left( \bar{x}_i - \frac{\Sigma w_i \bar{x}_i}{\Sigma w_i} \right) = 0,$$

we have

$$\frac{df(\lambda^2)}{d\lambda^2} = \sum_{i=1}^p \frac{dw_i}{d\lambda^2} \left( \bar{x}_i - \frac{\Sigma w_i \bar{x}_i}{\Sigma w_i} \right)^2 = \sum_{i=1}^p -w_i^2 \left( \bar{x}_i - \frac{\Sigma w_i \bar{x}_i}{\Sigma w_i} \right)^2 < 0,$$

which proves our statement.

Hence the lower confidence limit  $\lambda_1^2$  of  $\lambda^2$  is given by the root of the equation in  $\lambda^2$ :

$$(7) \quad F = \frac{N-p}{p-1} \frac{\sum_{j=1}^p \left\{ w_j \left( \bar{x}_j - \frac{\sum w_j \bar{x}_j}{\sum w_j} \right)^2 \right\}}{\sum \sum (x_{ij} - \bar{x}_j)^2}$$

and the upper confidence limit  $\lambda_2^2$  of  $\lambda^2$  is given by the root of the equation in  $\lambda^2$ :

$$(8) \quad F = F_1.$$

Since  $f(\lambda^2)$  is monotonically decreasing, the equations (7) and (8) have at most one root in  $\lambda^2$ . If the equation (7) or (8) has no root, the corresponding confidence limit has to be put equal to zero. If neither (7) nor (8) has a root, we have to reject at least one of the hypotheses:

- (1)  $x_{ij} = \epsilon_{ij} + \eta_j$ .
- (2) The variates  $\epsilon_{ij}$  and  $\eta_j$  ( $i = 1, \dots, m_j; j = 1, \dots, p$ ) are normally and independently distributed.
- (3) Each of the variates  $\epsilon_{ij}$  has the same distribution.
- (4) Each of the variates  $\eta_j$  has the same distribution.

The equations (7) and (8) are complicated algebraic equations in  $\lambda^2$ . For the actual calculation of the roots of these equations, well known approximation methods can be applied making use also of the fact that the left members are monotonic functions of  $\lambda^2$ . In applying any approximation method it is very useful to start with two limits of the root which do not lie far apart. We shall give here a method of finding such limits.

Denote by  $\bar{F}$  the function which we obtain from  $F$  (formula (6)) by substituting

$$\bar{w}_j = \frac{l_j}{1 + l_j \lambda^2} \text{ for } w_j \quad (j = 1, \dots, p).$$

Let  $\bar{f}$  be the function obtained from  $f$  by the same process.

Denote by  $\varphi(m, \lambda^2)$  the function which we obtain from  $\bar{F}$  by substituting  $m$  for  $l_1, \dots, l_p$ . We shall first show that  $\bar{F}$  is non-decreasing with increasing  $l_k$  ( $k = 1, \dots, p$ ), i.e.  $\frac{\partial \bar{F}}{\partial l_k} \geq 0$ . For this purpose we have only to show that

$\frac{\partial \bar{f}}{\partial l_k} \geq 0$ . We have:

$$\begin{aligned} \frac{\partial \bar{f}}{\partial l_k} &= \sum_j \frac{\partial \bar{w}_j}{\partial l_k} \left( \bar{x}_j - \frac{\sum \bar{w}_j \bar{x}_j}{\sum \bar{w}_j} \right)^2 - 2 \frac{\partial}{\partial l_k} \left( \frac{\sum \bar{w}_j \bar{x}_j}{\sum \bar{w}_j} \right) \cdot \left[ \sum \bar{w}_j \cdot \left( \bar{x}_j - \frac{\sum \bar{w}_j \bar{x}_j}{\sum \bar{w}_j} \right) \right] \\ &= \sum_j \frac{\partial \bar{w}_j}{\partial l_k} \left( \bar{x}_j - \frac{\sum \bar{w}_j \bar{x}_j}{\sum \bar{w}_j} \right)^2 = \frac{1}{(1 + l_k \lambda^2)^2} \left( \bar{x}_k - \frac{\sum \bar{w}_j \bar{x}_j}{\sum \bar{w}_j} \right)^2 \geq 0. \end{aligned}$$

Hence our statement is proved. Denote by  $m'$  the smallest and by  $m''$  the greatest of the values  $m_1, \dots, m_p$ . Then we obviously have

$$(9) \quad \varphi(m', \lambda^2) \leq F \leq \varphi(m'', \lambda^2).$$

Denote by  $\lambda_1'^2, \lambda_1''^2, \lambda_2'^2, \lambda_2''^2$  the roots in  $\lambda^2$  of the following equations respectively:

$$\varphi(m', \lambda^2) = F_2;$$

$$\varphi(m'', \lambda^2) = F_2;$$

$$\varphi(m', \lambda^2) = F_1; \quad \varphi(m'', \lambda^2) = F_1.$$

Since  $F$  is monotonically decreasing with increasing  $\lambda^2$ , on account of (7), (8), and (9) we obviously have

$$\lambda_1'^2 \leq \lambda_1^2 \leq \lambda_1''^2$$

and

$$\lambda_2'^2 \leq \lambda_2^2 \leq \lambda_2''^2.$$

The above inequalities give us the required limits.

COLUMBIA UNIVERSITY,  
NEW YORK, N. Y.

## THE DISTRIBUTION OF QUADRATIC FORMS IN NON-CENTRAL NORMAL RANDOM VARIABLES

BY WILLIAM G. MADOW<sup>1</sup>

The following theorem is the algebraic basis of the theorem of R. A. Fisher and W. G. Cochran which states necessary and sufficient conditions that a set of quadratic forms in normally and independently distributed random variables should themselves be independently distributed in  $\chi^2$ -distributions.<sup>2</sup>

**THEOREM I.** *If the real quadratic forms  $q_1, \dots, q_m$ , in  $x_1, \dots, x_n$ , are such that*

$$(1) \quad \sum_{\gamma} q_{\gamma} = \sum_{\nu} x_{\nu}^2,$$

*and if the rank of  $q_{\gamma}$  is  $n_{\gamma}$ , then a necessary and sufficient condition that*

$$(2) \quad q_{\gamma} = \sum z_{\alpha}^2,$$

<sup>1</sup> The letters  $i, j, \mu, \nu$  will assume all integral values from 1 through  $n$ , the letter  $\gamma$  will assume all integral values from 1 through  $m$ , ( $n \geq m$ ), the letter  $\alpha$  will assume all integral values from  $n_1 + \dots + n_{\gamma-1} + 1$  through  $n_1 + \dots + n_{\gamma}$ , ( $n_0 = 0, n_1 + \dots + n_m = n$ ), the letters  $\beta, \beta'$  will assume all integral values from 1 through  $n'$ , and the letters  $r, s$  will assume all integral values from 1 through  $n - 1$ .

<sup>2</sup> The references are: W. G. Cochran, "The Distribution of Quadratic Forms in a Normal System, with Applications to the Analysis of Covariance," *Proc. Camb. Phil. Soc.*, Vol. 30 (1934), pp. 178-191, and R. A. Fisher, "Applications of 'Student's' Distribution," *Metron*, Vol. 5 (1926), pp. 90-104.

where the real linear functions  $z_\beta$  of the  $x_r$  are defined by

$$(3) \quad x_r = \sum_{\beta} c_{r\beta} z_{\beta}$$

is

$$(4) \quad n' = n.$$

Furthermore the system of linear forms (3) constitute an orthogonal transformation.

**PROOF: Necessity.** Since the rank of a sum of quadratic forms is less than or equal to the sum of their ranks, it follows that  $n' \geq n$ . Upon substituting from (3) for the  $x$ 's in (1), and using (2), it is seen that, for all values of the  $z$ 's,

$$\sum_{\beta} z_{\beta}^2 = \sum_{\beta, \beta'} \left( \sum_r c_{r\beta} c_{r\beta'} \right) z_{\beta} z_{\beta'},$$

and hence, from (1), it follows that

$$(5) \quad \sum_r c_{r\beta} c_{r\beta'} = \delta_{\beta\beta'},$$

where  $\delta_{\beta\beta'} = 0$ , if  $\beta \neq \beta'$ , and  $\delta_{\beta\beta'} = 1$  if  $\beta = \beta'$ . However, since the rank of the system of linear forms (3) is not greater than  $n$ , and since the matrix of (5) is the product of the matrix of (3) by its transposed matrix, it follows that (5) can be true only if  $n'$  is not greater than  $n$ . Consequently  $n' = n$ . It then is an immediate result of (5) that the transformation (3) is orthogonal.

**Sufficiency.** We assume that  $n' = n$ . By a real linear transformation of  $x_1, \dots, x_n$  we obtain linear forms  $z_r$  such that

$$q_r = \sum_{\alpha} c_{\alpha} z_{\alpha}^2,$$

where  $c_{\alpha} = 1$  or  $-1$ . The set of linear functions  $z_1, \dots, z_n$  are linearly independent, for if  $z_n \neq 0$ , and if real numbers  $h_1, \dots, h_{n-1}$  not all zero, exist such that, say,

$$z_n = \sum_r h_r z_r$$

then

$$\sum_r z_r^2 = \sum_{r,s} H_{rs} z_r z_s.$$

Substituting, we have

$$\sum_r q_r = \sum_r c_r z_r^2 = \sum_{r,s} \sum_{\mu,\nu} H_{rs} c_r^{\mu} c_s^{\nu} x_{\mu} x_{\nu},$$

where  $z_r = \sum_{\mu} c_r^{\mu} x_{\mu}$ . (It is not assumed here that the matrix of the  $c^{\mu}$  is the inverse of the matrix of the  $c_{\mu}$ . That fact is a consequence of this proof.)

Denoting the matrix of  $z_1, \dots, z_{n-1}$  by  $\bar{C}_n$  we see that the matrix of  $\sum_r q_r$  is  $\bar{C}_n' H \bar{C}_n$  where  $H$  is the matrix of the  $H_{rs}$  and has rank less than or equal to  $n - 1$  which contradicts the hypothesis. Hence if  $C$  is the matrix having the elements

$c$ , in its main diagonal and zeros elsewhere and if  $C_n$  is the matrix of  $z_1, \dots, z_n$  it follows that

$$C'_n C C_n = I,$$

where  $I$  is the identity matrix, i.e. the matrix having ones in the main diagonal and zeros elsewhere and  $C_n$  non-singular. Then  $C = C_n^{-1}/C_n^{-1}$  and hence  $C$  is the identity matrix and  $C_n$  is orthogonal.

Among the hypotheses of the Fisher-Cochran theorem is the hypothesis that the mean value of  $x_\mu$  is 0, and the variance of  $x_\mu$  is  $\sigma^2$ . However, in connection with his analysis of the distribution of the multiple correlation coefficient,<sup>3</sup> R. A. Fisher derived the distribution of the sum of the squares of  $n$  independently distributed random variables  $x_1, \dots, x_n$ , the probability density of  $x_\mu$  being given by

$$(6) \quad p(x_\mu) = (2\pi\sigma^2)^{-1} \exp \left[ -\frac{1}{2\sigma^2} (x_\mu - a_\mu)^2 \right].$$

More recently, P. C. Tang,<sup>4</sup> has used the distribution of the sum of non-central squares in his study of the power function of the analysis of variance test.

In this note we extend the Fisher-Cochran theorem to non-central random variables. If the random variables  $x_\mu$  are independently distributed with probability densities given by (6), Fisher and Tang have shown that if  $\chi'^2 = \frac{1}{\sigma^2} \sum_\nu x_\nu^2$ , then the probability density of  $\chi'^2$  is given by

$$(7) \quad p(\chi'^2) = \frac{1}{2} e^{-\lambda} (\frac{1}{2} \chi'^2)^{\frac{1}{2}n-1} e^{-\frac{1}{2} \chi'^2} \sum_{\nu=0}^{\infty} \frac{(\frac{1}{2} \lambda \chi'^2)^\nu}{\nu! \Gamma(\frac{1}{2}n + \nu)}$$

where  $\lambda = \frac{1}{2\sigma^2} \sum_\nu a_\nu^2$ .

We now give necessary and sufficient conditions that a set of quadratic forms in normally and independently distributed random variables should themselves be independently distributed in  $\chi'^2$ -distributions.

**THEOREM II.** Let  $x_1, \dots, x_n$  be independently distributed random variables, the random variable  $x_\mu$  having probability density (6). Denote  $\sum_\nu x_\nu^2$  by  $q$ , and

denote  $\frac{1}{2\sigma^2} \sum_\nu a_\nu^2$  by  $\lambda$ . Let  $q_1, \dots, q_m$ , be quadratic forms,

$$q_\gamma = \sum_{\mu, \nu} a_{\mu\nu}^{(\gamma)} x_\mu x_\nu$$

such that  $\sum_\gamma q_\gamma = q$ , and let the rank of  $q_\gamma$  be denoted by  $n_\gamma$ .

<sup>3</sup> R. A. Fisher, "The General Sampling Distribution of the Multiple Correlation Coefficient," *Proc. Royal Soc. of London, (A)*, Vol. 121 (1928), pp. 654-673.

<sup>4</sup> P. C. Tang, "The Power Function of the Analysis of Variance Tests with Tables and Illustrations of their Use," *Statistical Research Memoirs*, Vol. 2 (1938), pp. 126-149.

A necessary and sufficient condition that the quadratic forms  $\chi'_\gamma$ ,  $\left(\chi'_\gamma = \frac{q_\gamma}{\sigma^2}\right)$ , be independently distributed with joint probability density

$$(8) \quad p(\chi_1'^2, \dots, \chi_m'^2) = \prod_\gamma p(\chi_\gamma'^2),$$

where  $p(\chi_\gamma'^2)$  is given by (7) with  $n_\gamma$  and  $\lambda_\gamma$  in place of  $n$  and  $\lambda$ , and

$$(9) \quad \lambda_\gamma = \frac{1}{2\sigma^2} \sum_{\mu, \nu} a_{\mu\nu}^{(\gamma)} a_\mu a_\nu$$

is  $n' = n$ .

PROOF. *Necessity.* Tang<sup>5</sup> has shown that the distribution of  $\chi'^2$  is given by (7) and that if the  $\chi'_\gamma$  have joint distribution (8), then the distribution of  $\chi_1'^2 + \dots + \chi_m'^2$ , ( $= \chi'^2$ ), is (7) with  $n'$  in place of  $n$ . Upon comparing terms, we see that  $n' = n$ .

*Sufficiency.* By Theorem I there exist  $n$  orthogonal linear functions (3) such that (2) is true. Then it is easy to see that the random variables  $z_1, \dots, z_n$  are independently distributed with a joint probability density

$$(10) \quad p(z_1, \dots, z_n) = (2\pi\sigma^2)^{-1/2n} \exp \left[ -\frac{1}{2} \sum_\gamma (z_\gamma - a'_\gamma)^2 \right],$$

where

$$\sum_\gamma a_\gamma'^2 = \sum_\gamma a_\gamma^2, \quad \text{and} \quad a'_\mu = \sum_\gamma c_{\mu\gamma} a_\gamma.$$

If we set  $2\sigma^2\lambda_\gamma = \sum_\alpha a_\alpha'^2$ , then we have, from (7) and (10), that the  $\chi'_\gamma$  are independently distributed with joint probability density (8). It is only necessary to show that  $\sum_\alpha a_\alpha'^2 = \sum_{\mu, \nu} a_{\mu\nu}^{(\gamma)} a_\mu a_\nu$ , in order to complete the proof of the theorem. Now

$$\sum_{\mu, \nu} a_{\mu\nu}^{(\gamma)} a_\mu a_\nu = \sum_{i, j} \left( \sum_{\mu, \nu} a_{\mu\nu}^{(\gamma)} c_{i\mu} c_{j\nu} \right) a'_i a'_j.$$

On the other hand, by direct substitution for the  $z$ 's we see that

$$q_\gamma = \sum_\alpha z_\alpha^2 = \sum_{\mu, \nu} \left( \sum_\alpha c_{\mu\alpha} c_{\nu\alpha} \right) x_\mu x_\nu$$

and hence  $a_{\mu\nu}^{(\gamma)} = \sum_\alpha c_{\mu\alpha} c_{\nu\alpha}$ . Since (1) is an orthogonal transformation,

$$\sum_{\mu, \nu} a_{\mu\nu}^{(\gamma)} c_{i\mu} c_{j\nu} = \sum_{\mu, \nu} \left( \sum_\alpha c_{\mu\alpha} c_{\nu\alpha} \right) c_{i\mu} c_{j\nu} = \sum_\alpha \delta_{\alpha i} \delta_{\alpha j},$$

where  $\delta_{\alpha i} = 0$ , if  $\alpha \neq i$  and  $= 1$  if  $\alpha = i$ , which completes the proof.

It is emphasized that the form of  $\lambda_\gamma$  makes it unnecessary to calculate the matrix of  $q_\gamma$  to determine  $\lambda_\gamma$  since the values  $a_\nu$  need only be substituted for the  $x_\nu$  in the original expression for  $q_\gamma$  to determine  $\lambda_\gamma$ .

WASHINGTON, D. C.

<sup>5</sup> See 4 p. 140.

## TWO PROPERTIES OF SUFFICIENT STATISTICS

BY LOUIS OLSHEVSKY

The concept of sufficient statistics was introduced by R. A. Fisher in 1922. It was refined and extended in 1936 by Neyman and Pearson who gave definitions of shared sufficient statistics and sufficient sets of algebraically independent statistics.<sup>1</sup> Today the concept plays an important part in the theory of the subject. Characterized briefly, a statistic associated with a single or specific population parameter is sufficient when no other statistic calculated from the same sample sheds any additional light on the value of the parameter. We shall prove that sets of sufficient statistics possess certain interconnections so that when one set is known every other set with a like number of members and linked with the same population parameters is discoverable.

**THEOREM 1.** *If  $T_1, \dots, T_m$  are a set of  $m$  ( $m \leq n$ ) algebraically independent sufficient statistics with regard to the parameters  $\theta_1, \dots, \theta_q$  and the probability law  $p(x_1, \dots, x_n | \theta_1, \dots, \theta_q, \dots, \theta_l)$ , a necessary and sufficient condition for the sufficiency of any set of  $m$  algebraically independent statistics  $T'_1, \dots, T'_m$  with regard to the same parameters and the same probability distribution is that the  $T'_i$  be a set of independent functions of the  $T_i$  ( $i, j = 1, \dots, m$ ).*

**PROOF:** As an adjunct in the demonstration we cite the following theorem due to Neyman.<sup>2</sup> For a set of algebraically independent statistics  $T_1, \dots, T_m$  to be a sufficient set with regard to the parameters  $\theta_1, \dots, \theta_q$ , it is necessary and sufficient that in any point of sample space, except perhaps for a set of measure zero, it should be possible to present the probability law in the form of the product

$$(1) \quad p(x_1, \dots, x_n | \theta_1, \dots, \theta_q, \dots, \theta_l) \\ = p(T_1, \dots, T_m | \theta_1, \dots, \theta_q) \cdot \phi(x_1, \dots, x_n; \theta_{q+1}, \dots, \theta_l)$$

where  $p(T_1, \dots, T_m | \theta_1, \dots, \theta_q)$  is the probability law of  $T_1, \dots, T_m$  and the function  $\phi$  does not depend upon  $\theta_1, \dots, \theta_q$ .

The sufficiency of the condition stated in the hypothesis of Theorem I is now immediately evident. For, if  $p'$  and  $\phi'$  refer to the second set of algebraically independent statistics and  $T'_i = T'_i(T_1, \dots, T_m)$  where the functions are independent, the relations can be solved for the  $T_i$  in terms of the  $T'_i$  giving  $T_i = T_i(T'_1, \dots, T'_m)$ ,  $p'(T'_1, \dots, T'_m | \theta_1, \dots, \theta_q)$

$$= p[T_1(T'_1, \dots, T'_m), \dots, T_m(T'_1, \dots, T'_m) | \theta_1, \dots, \theta_q] \frac{\partial(T_1, \dots, T_m)}{\partial(T'_1, \dots, T'_m)}, \\ \phi'(x_1, \dots, x_n; \theta_{q+1}, \dots, \theta_l) = \phi(x_1, \dots, x_n; \theta_{q+1}, \dots, \theta_l) \div \frac{\partial(T_1, \dots, T_m)}{\partial(T'_1, \dots, T'_m)},$$

<sup>1</sup> See Neyman and Pearson: "Sufficient Statistics and Uniformly Most Powerful Tests of Statistical Hypotheses," *Statistical Research Memoirs of the University of London*, June 1936. The notation of the present paper is taken from this article.

<sup>2</sup> See Neyman's article in the *Giornale dell' Istituto Italiano degli Attuari*, Vol. VI, No. 4 (1935) as well as the memoir referred to in footnote 1.

and

$$(2) \quad p(x_1, \dots, x_n | \theta_1, \dots, \theta_q, \dots, \theta_l) \\ = p'(T'_1, \dots, T'_m | \theta_1, \dots, \theta_q) \cdot \phi'(x_1, \dots, x_n; \theta_{q+1}, \dots, \theta_l).$$

Proof of the necessity is somewhat more involved. Since the  $T_i$  and  $T'_i$  are both sets of algebraically independent statistics with regard to  $\theta_1, \dots, \theta_q$ , equations (1) and (2) are satisfied. They are, in fact, identities when the values of  $T_1, \dots, T_m$  and  $T'_1, \dots, T'_m$  in terms of the  $x_i$  are substituted. Division of (1) by (2) and multiplication leads to the equation

$$(3) \quad \frac{p(T_1, \dots, T_m | \theta_1, \dots, \theta_q)}{p'(T'_1, \dots, T'_m | \theta_1, \dots, \theta_q)} = \frac{\phi'(x_1, \dots, x_n; \theta_{q+1}, \dots, \theta_l)}{\phi(x_1, \dots, x_n; \theta_{q+1}, \dots, \theta_l)}.$$

The right side of (3) is free of  $\theta_1, \dots, \theta_q$ . Therefore, in reality the left side must be too. If some or all of the parameters  $\theta_1, \dots, \theta_q$  enter formally into the left side, we can choose  $m+1$  sets of values  $\theta_1^i, \dots, \theta_q^i$  ( $i = 1, \dots, m+1$ ) such that each of the  $m+1$  functions  $p(T_1, \dots, T_m | \theta_1^i, \dots, \theta_q^i) \div p'(T'_1, \dots, T'_m | \theta_1^i, \dots, \theta_q^i)$  differs formally from all of the others. We can, then, since each is equal to the right side of (3) which is free of  $\theta_1, \dots, \theta_q$ , equate any one of these functions to the remaining  $m$  in turn. This provides  $m$  independent equations whose very existence proves that the  $T'_i$  are functions of the  $T_i$  and vice versa.

If none of the parameters  $\theta_1, \dots, \theta_q$  enters formally into the left side of (3),  $p(T_1, \dots, T_m | \theta_1, \dots, \theta_q)$  must be of the form  $p(T_1, \dots, T_m)g(\theta_1, \dots, \theta_q)$  and  $p'(T'_1, \dots, T'_m | \theta_1, \dots, \theta_q)$  of the form  $p'(T'_1, \dots, T'_m)g(\theta_1, \dots, \theta_q)$ . In this case the original probability law  $p(x_1, \dots, x_n | \theta_1, \dots, \theta_q, \dots, \theta_l)$  contains  $\theta_1, \dots, \theta_q$  only nominally and there can be no talk of any statistics designed to estimate these parameters either singly or in combination.

When  $m = 1$  and the set of algebraically independent statistics reduces to one, the single statistic is termed a shared sufficient statistic of the parameters  $\theta_1, \dots, \theta_q$ .<sup>8</sup> For this special case, Theorem I can be restated as follows. If  $T$  is a shared sufficient statistic with regard to the population parameters  $\theta_1, \dots, \theta_q$  and the probability distribution  $p(x_1, \dots, x_n | \theta_1, \dots, \theta_q, \dots, \theta_l)$ , the necessary and sufficient condition for the sufficiency of any statistic  $T'$  with regard to the same parameters and the same probability distribution is that  $T'$  be a function of  $T$ . When  $m$  and  $q$  both equal one, the statistic becomes a sufficient statistic in the sense originally defined by Fisher in 1922.

A physical law is independent of the coordinate system used to express it. This fact is taken account of in modern physics through the employment of tensors. One might hope for a parallel situation in the relation between sufficient statistics and the probability law to which they refer. Given any  $l$  parameter family of distribution laws  $p(x_1, \dots, x_n | \theta_1, \dots, \theta_l)$ , the substitu-

<sup>8</sup> See the memoir mentioned in footnote 1.



tion  $\theta_i = \theta_i(\theta'_1, \dots, \theta'_l)$  ( $i = 1, \dots, l$ ) leads to the equally valid representation of the family

$$p'(x_1, \dots, x_n | \theta'_1, \dots, \theta'_l) \\ = p[x_1, \dots, x_n | \theta_1(\theta'_1, \dots, \theta'_l), \dots, \theta_l(\theta'_1, \dots, \theta'_l)].$$

Is a set of statistics sufficient with respect to the first representation also sufficient with respect to the second? The answer is partly in the affirmative and is given by the following proposition.

**THEOREM II.** *If the set of algebraically independent statistics  $T_1, \dots, T_m$  is sufficient with regard to the parameters  $\theta_1, \dots, \theta_q$  and the probability law  $p(x_1, \dots, x_n | \theta_1, \dots, \theta_q, \dots, \theta_l)$ , it is also sufficient with regard to  $\theta'_1, \dots, \theta'_q$  and any other representation  $p'(x_1, \dots, x_n | \theta'_1, \dots, \theta'_q, \dots, \theta'_l)$  of the same probability law provided  $\theta'_i$  ( $i = 1, \dots, q$ ) are independent functions of  $\theta_1, \dots, \theta_q$  only and  $\theta'_j$  ( $j = q + 1, \dots, l$ ) are functions of  $\theta_{q+1}, \dots, \theta_l$  only.*

**PROOF:** The proof of the theorem is obvious. We are given the fact that  $p(x_1, \dots, x_n | \theta_1, \dots, \theta_q, \dots, \theta_l) = p(T_1, \dots, T_m | \theta_1, \dots, \theta_q) \cdot \phi(x_1, \dots, x_n; \theta_{q+1}, \dots, \theta_l)$ . Since the  $\theta'_i$  ( $i = 1, \dots, q$ ) are functions of  $\theta_1, \dots, \theta_q$  only and the  $\theta'_j$  ( $j = q + 1, \dots, l$ ) are functions of  $\theta_{q+1}, \dots, \theta_l$  only, it follows that  $\theta_i = \theta_i(\theta'_1, \dots, \theta'_q)$  ( $i = 1, \dots, q$ ) and  $\theta_j = \theta_j(\theta'_{q+1}, \dots, \theta'_l)$  ( $j = q + 1, \dots, l$ ). Consequently,

$$(4) \quad p'(x_1, \dots, x_n | \theta'_1, \dots, \theta'_q, \dots, \theta'_l) \\ = p'(T_1, \dots, T_m | \theta'_1, \dots, \theta'_q) \cdot \phi'(x_1, \dots, x_n; \theta'_{q+1}, \dots, \theta'_l)$$

and the theorem is established.

NEW YORK, N. Y.

## NOTE ON THE MOMENTS OF A BINOMIALLY DISTRIBUTED VARIATE

BY W. D. EVANS

J. A. Joseph, has given two interesting triangular arrangements of numbers, the second of which is reproduced herewith as Table 1.<sup>1</sup> The successive rows in this table are the coefficients in the expansion of  $x^n$  as a function of the factorials  $x^{(i)}$ , using the notation of the calculus of finite differences. For example,

$$x^4 = x^{(4)} + 6x^{(3)} + 7x^{(2)} + x,$$

where

$$x^{(i)} = x(x-1)(x-2) \dots (x-i+1).$$

Joseph points out that the coefficients may be used to generate the numbers of Laplace.

<sup>1</sup> J. A. Joseph, "On the Coefficients of the Expansion of  $X^{(n)}$ ," *Annals of Math. Stat.*, Vol. X (1939), p. 293.

A general expression defining any of the coefficients in terms of its place of occurrence in Table 1 may be set up. If we denote by  $F_c(r)$  the number in row  $r$  and column  $c$  of the table, we have

$$(1) \quad F_c(r) = \sum_{k_1=0}^{r-c+1} k_1 \sum_{k_2=0}^{k_1} k_2 \sum_{k_3=0}^{k_2} k_3 \cdots \sum_{k_{c-1}=0}^{k_{c-2}} k_{c-1} \quad (r \geq c).$$

This expression is of additional interest since the numbers defined by it are likewise the coefficients in the expression of the  $z$ -th moment about the origin of a binomially distributed variate in terms of the probability of the variate and the size of the sample in which it is contained. For example, it may be easily

TABLE 1

	1	2	3	4	5	...	$c$
1	1						
2	1	1					
3	1	3	1				
4	1	6	7	1			
5	1	10	25	15	1		
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
	$F_1(r)$	$F_2(r)$	$F_3(r)$	$F_4(r)$	$F_5(r)$	...	$F_c(r)$

verified that if  $\alpha$  is such a variate,  $p$  its probability of occurrence, and  $n$  the size of the sample in which it is contained,

$$E(\alpha)^2 = n^{(2)}p^2 + np$$

$$E(\alpha)^3 = n^{(3)}p^3 + 3n^{(2)}p^2 + np$$

$$E(\alpha)^4 = n^{(4)}p^4 + 6n^{(3)}p^3 + 7n^{(2)}p^2 + np$$

and so on.

Ordinarily, computation of the higher moments of a binomially distributed variate is a tedious process of repeated differentiation. However, equation (1) immediately permits us to generalize the foregoing expressions to give the  $z$ -th moment of  $\alpha$  as follows:

$$(2) \quad E(\alpha)^z = \sum_{i=0}^{z-1} n^{(z-i)} p^{z-i} \sum_{k_1=0}^{i-1} k_1 \sum_{k_2=0}^{k_1} k_2 \cdots \sum_{k_{i-1}=0}^{k_{i-2}} k_{i-1}.$$

It will be noted that when  $c = 1$  in equation (1) and  $i$  in equation (2) are equal to zero, the repeated summations vanish to be replaced by the value one.

By means of equation (2) much of the labor usually involved in expressing the  $z$ -th moment about the origin of a binomially distributed variate in terms of  $n$  and  $p$  may be avoided.

## REPORT OF THE ANNUAL MEETING OF THE INSTITUTE

The fifth annual meeting of the Institute of Mathematical Statistics was held in Philadelphia, Pennsylvania, on December 27 and 28, 1939, in conjunction with the meetings of the American Statistical Association, the Econometric Society, and the American Sociological Society. The program for the meeting was arranged by Professor C. C. Craig.

On Wednesday morning, December 27, the Institute held a session devoted to contributed papers on *Statistical Theory* and *Methodology*. Professor P. R. Rider, President of the Institute, presided. At that time the following papers were presented:

1. *On the unbiased character of certain likelihood-ratio tests when applied to normal systems.*  
Joseph F. Daly, The Catholic University of America.
2. *The product seminvariants of the mean and a central moment in samples.*  
C. C. Craig, University of Michigan.
3. *A method for minimizing the sum of absolute values of deviations.*  
Robert Singleton, Princeton Local Government Survey.
4. *On certain criteria for testing the homogeneity of  $k$  estimates of variance.*  
C. Eisenhart and Frieda S. Swed, University of Wisconsin.
5. *On a test whether two samples are from the same population.*  
A. Wald and J. Wolfowitz, Columbia University and Brooklyn, New York.
6. *The power functions of certain tests of significance in harmonic analysis and lag correlation.*  
William G. Madow, Washington, D. C.
7. *Some theoretical aspects of the use of transformations in the statistical analysis of replicated experiments.*  
W. G. Cochran, Iowa State College.
8. *The standard errors of geometric and harmonic types of index numbers.*  
Nilan Norris, Hunter College.
9. *A study of R. A. Fisher's  $z$  distribution and the related  $F$  distribution.*  
L. A. Aroian, Hunter College.
10. *A note on the analysis of variance with unequal class frequencies.*  
Abraham Wald, Columbia University.
11. *An approach to problems involving disproportionate frequencies.*  
Burton D. Seeley, U. S. Department of Labor.

Abstracts of these papers are given at the close of this report.

Immediately following the session just described, the Institute held its annual business meeting. At that time President Rider announced that the newly elected officers for the year 1940 are: President, S. S. Wilks, Princeton University; Vice-Presidents: C. C. Craig, University of Michigan, and A. T. Craig, University of Iowa; Secretary-Treasurer: P. R. Rider, Washington University.

At one o'clock on the same day, members of the Institute and their guests

attended the annual luncheon. At the luncheon, Professor B. H. Camp addressed the Institute on *Non-standard Deviations*.

On Wednesday afternoon, the Institute met jointly with the American Statistical Association for a program devoted to *Lag Effects in Statistics and Economics*. Professor J. D. Tamarkin presided and at this time the following papers were read:

1. *Lag effects in statistics and related problems.*  
A. J. Lotka, Metropolitan Life Insurance Company.
2. *Some methods in the analysis of lag effects.*  
H. T. Davis, Northwestern University.
3. *Lag effects in economics.*  
Charles F. Roos, Institute of Applied Econometrics, Inc.

A joint session with the Biometric Section of the American Statistical Association was held on Wednesday evening, Professor George W. Snedecor presiding. The papers presented at this session, which dealt with *Design and Analysis of Replicated Experiments*, were the following:

1. *Practical difficulties met in the use of experimental designs.*  
A. E. Brandt, Soil Conservation Service.
2. *Factorial design and covariance in the biological assay of vitamin D.*  
C. I. Bliss, Sandusky, Ohio.
3. *Combinatorial problems in the design of experiments.*  
Gertude M. Cox, Iowa State College.
4. *Experimental trials with balanced incomplete blocks.*  
W. J. Youden, Boyce Thompson Institute.

On Thursday afternoon the Institute held consecutively joint sessions with the American Sociological Society and the Econometric Society. At the first of these, Professor William F. Ogburn presided and the following program was presented:

1. *How the mathematician can help the sociologist.*  
Samuel A. Stouffer, University of Chicago.
  2. *Some problems of combinations and permutations as they apply to a comprehensive classification of social groups.*  
George A. Lundberg, Bennington College.
- Discussion: C. C. Craig, University of Michigan.  
Philip M. Houser, U. S. Bureau of the Census.

At the second session the topic for discussion was *Recent Advances in Business Cycle Analysis* and these papers were given:

1. *Recursive methods in business cycle analysis.*  
Merrill M. Flood, Princeton Surveys.
2. *An appreciation of some recent mathematical business cycle theories.*  
Gerhard Tintner, Iowa State College.
3. *The statisticians' new clothiers.*  
Arne Fisher, Western Union Telegraph Company.

PAUL R. RIDER, *Secretary*.

## ABSTRACTS OF PAPERS

(Presented on December 27, 1939, at the Philadelphia meeting of the Institute)

### On the Unbiased Character of Certain Likelihood-Ratio Tests when Applied to Normal Systems. JOSEPH F. DALY, The Catholic University of America.

Consider a random sample of  $N$  observations on a set of variates  $x^1, \dots, x^q$ , where  $x^1, \dots, x^k$  are assumed to be normally distributed about means which are linear functions  $m^i = \Sigma b_{\mu}^i x^{\mu}$  of the fixed variates  $x^{k+1}, \dots, x^q$ . One is sometimes required to decide whether the sample tends to contradict the further hypothesis,  $H_0$ , that the coefficients  $b_{\mu}^i$  belonging to a certain subset of the fixed variates, say  $x^{k+1}, \dots, x^{k+h}$ , have the specific values  $b_{\mu_0}^i$ . Such a situation occurs, for example, in the generalized analysis of variance. In this paper it is shown that the Neyman-Pearson method of the ratio of likelihoods yields a test of  $H_0$  which is (at least locally) unbiased; in other words, this test is less likely to reject  $H_0$  when the sample is in fact drawn from a normal population in which  $b_{\mu}^i = b_{\mu_0}^i$  than when it is drawn from a normal population in which the  $b_{\mu}^i$  are different from but sufficiently close to  $b_{\mu_0}^i$ . In the special cases  $k = 1$  or  $h = 1$  the proof goes through even without the restriction that the true  $b_{\mu}^i$  be close to  $b_{\mu_0}^i$ , a result which is also implicit in the papers by P. C. Tang and P. L. Hsu (*Stat. Res. Mem.* Vol. 2).

Similarly with respect to the hypothesis  $H_1$  that the deviations  $x^i - \Sigma b_{\mu}^i x^{\mu}$  fall into certain mutually independent sets the  $\lambda$ -test is at least locally unbiased; and it has the additional property that the expected value of any positive integral power of  $\sqrt{N/\lambda}$  is greater when  $H_1$  is true than when the sample is drawn from any other normal population.

### The Product Seminvariants of the Mean and a Central Moment in Samples. C. C. CRAIG, The University of Michigan.

The method used by the author in calculating the product seminvariants of a pair of central moments in samples is not adapted without modification to the present problem. In the present paper the necessary modification is developed which gives a routine method for the calculation of these sampling distribution characteristics. The calculation is a little heavier than in the previous case but the results for the mean and the second, third, and fourth central moments are given up to the fourth order except in one case in which the weight is 13. It is planned to follow this with a further study of the distribution of Fisher's  $t$  in samples from a normal population.

### A Method for Minimizing the Sum of Absolute Values of Deviations. ROBERT SINGLETON, Princeton Local Government Survey.

E. C. Rhodes (*Philosophical Magazine*, May 1930) presented a method for the estimation of parameters in a linear regression where it is desired to minimize the sum of absolute values of the deviations. In this paper the structure of the deviation surface is analyzed and a method of steepest descent is developed which for computational purposes is an improvement over Rhodes' method. The process is finite and leads to an exact solution. The method and the formulae used are such as to permit the successive additions of new observations or sets of observations to the original data, or the exclusion of an observation from the original set, and the determination of the parameters for the sets of data so derived, with little additional labor.

**On Certain Criteria for Testing the Homogeneity of  $k$  Estimates of Variance.**  
C. EISENHART AND FRIEDA S. SWED, University of Wisconsin.

Given  $k$  variance estimates  $s_1^2, s_2^2, \dots, s_k^2$  with  $n_r s_r^2$ , ( $r = 1, 2, \dots, k$ ), independently distributed as  $\chi^2 \sigma^2$  for  $n_r$  degrees of freedom, tests of the hypothesis,  $H_0$ , that  $\sigma_r^2 = \sigma^2$ , ( $r = 1, 2, \dots, k$ ), where  $\sigma^2$  is unknown, have been based to date on one or the other of the quantities

$$Q_1 = \sum_{r=1}^k n_r (s_r^2 - \bar{s}^2)^2 / 2\bar{s}^4$$

$$Q_2 = w \log (n\bar{s}^2/w) - \sum_{r=1}^k w_r \log \{n_r s_r^2 / w_r\}$$

where the  $w_r$  are weights,  $w = \sum_{r=1}^k w_r$ ,  $n = \sum_{r=1}^k n_r$ , and  $n\bar{s}^2 = \sum_{r=1}^k n_r s_r^2$ . A. E. Brandt and

W. L. Stevens have advocated the use of  $Q_1$ , referring an observed value of  $Q_1$  to the  $\chi^2$  distribution for  $k - 1$  degrees of freedom. J. Neyman, E. S. Pearson, B. L. Welch, and M. S. Bartlett have advocated tests based on  $Q_2$ , Bartlett definitely proposing the use of degrees of freedom as weights, i.e.  $w_r = n_r$ , and recent work of E. J. G. Pitman and others has shown that unless  $w_r = n_r$  tests based on  $Q_2$  are biased. (A statistical test of an hypothesis  $H$  is said to be unbiased when the probability of rejecting  $H$  by its use is a minimum when  $H$  is true; obviously a desirable property.) When  $w_r = n_r$ , Bartlett has suggested that

the distribution of  $Q_2$  can be satisfactorily approximated by referring  $Q_2 / \left\{ 1 + \frac{1}{3(k-1)} \left( \sum_{r=1}^k \frac{1}{n_r} - \frac{1}{n} \right) \right\}$  to the  $\chi^2$  distribution for  $k - 1$  degrees of freedom. In this paper we discuss

the adequacy of the  $\chi^2$  distribution to describe the distribution of  $Q_1$  and of the adjusted  $Q_2$  when the degrees of freedom,  $n_r$ , are small.

U. S. Nair and D. J. Bishop have given theoretical evidence which suggests that when  $n_r \geq 2$ , ( $r = 1, 2, \dots, k$ ), Bartlett's adjusted  $Q_2$  may be expected to conform to the  $\chi^2$  distribution reasonably well in the neighborhood of the 5% and 1% levels. Using 1000 samples of 4 for which  $n_r s_r^2 / (n_{r+1})$  has been tabulated by W. A. Shewhart in Table D, Appendix II of his "Economic Control of Quality of Manufactured Product," 200 values of  $Q_1$  and  $Q_2$  (with adjustment) were calculated and compared with the  $\chi^2$  distribution for  $k - 1$  degrees of freedom. Two cases were studied: Case I,  $k = 5$  and  $n_1 = n_2 = \dots = 3$ ; Case II,  $k = 3$  and  $n_1 = n_2 = 3$  while  $n_3 = 9$ . As measured by the Chi-Square Goodness of Fit Test, using 11 degrees of freedom, the fits were good in all four instances. In Case I, for Bartlett's adjusted  $Q_2$  the test led to  $.80 < P < .90$ , and to  $.70 < P < .80$  for the Brandt-Stevens  $Q_1$ ; in Case II, the fits were poorer with  $.50 < P < .70$  for Bartlett's criterion and  $.10 < P < .20$  for the Brandt-Stevens. However, an examination of the *descending* cumulative distributions showed that in all instances these criteria exhibited a deficiency of large values of  $\chi^2$ , with the deficiency, in general, more marked in the case of the Brandt-Stevens test. Consequently, when one uses significance levels for these criteria obtained by means of the  $\chi^2$  approximation advocated, one is in reality using a level of significance slightly less than that professed. The discrepancy is not great, however, and is on the safe side, i.e. one will reject  $H_0$  falsely in the long run less often than one professes to be doing. Without doubt, however, one will also detect the falsehood of  $H_0$  when  $\sigma_r^2 \neq \sigma_t^2$ , for at least one pair of values of  $r$  and  $t$ ,  $r \neq t$ , less often in the long run by the use of these approximate significance levels than if the true levels were used, but we have no definite evidence at present on this point. A somewhat disquieting feature is that the agreement between the  $\chi^2$  values yielded by the two criteria becomes worse as one proceeds toward larger values of  $\chi^2$  in

terms of either quantity. Thus, of 8 samples which  $Q_2$  would have rejected at the 5% level in Case I, only 4 of these would have been rejected by  $Q_1$ , and  $Q_2$  would have passed 3 samples of the 7 rejected by  $Q_1$ . Thus it appears that, if one wishes to work with a given chance of rejecting  $H_0$  falsely, one should choose one of these criteria and then stick to it in future applications. For large values of the  $n_r$  the two criteria tend to equivalence, so the choice between them is of interest mainly for small  $n_r$ , but cannot be made with full information until more is known about the bias, if any, of the Brandt-Stevens test, and the relative power of the two tests with regard to alternatives to  $H_0$ .

**On a Test Whether Two Samples are from the Same Population.** A. WALD AND J. WOLFOWITZ, Columbia University and Brooklyn, New York.

Let  $X$  and  $Y$  be two independent random variables about whose distributions nothing is known except that they are continuous. Let  $x_1, x_2, \dots, x_m$  be a set of  $m$  independent observations on  $X$  and let  $y_1, y_2, \dots, y_n$  be a set of  $n$  independent observations on  $Y$ . The null hypothesis to be tested is that the distributions of  $X$  and  $Y$  are identical.

Let the set of  $m + n$  observations be arranged in order of magnitude, thus:  $z_1, z_2, \dots, z_{m+n}$ . Replace  $z_i$  by  $v_i$  ( $i = 1, 2, \dots, m + n$ ) where  $v_i = 0$  if  $z_i$  is a member of the set of  $x$ 's and  $v_i = 1$ , if  $z_i$  is a member of the set of  $y$ 's. Since the null hypothesis states only that the distributions of  $X$  and  $Y$  are identical without specifying them in any other way, the distribution of the statistic  $U$  used for testing the null hypothesis must be independent of this common distribution of  $X$  and  $Y$ . It can easily be shown that the statistic  $U$  must be a function only of the sequence  $v_1, v_2, \dots, v_{m+n}$ .

A subsequence  $v_s, v_{s+1}, \dots, v_{s+r}$  (where  $r$  may also be 0) is called a run if  $v_s = v_{s+1} = \dots = v_{s+r}$  and if  $v_{s-1} \neq v_s$  when  $s < 1$  and if  $v_{s+r} \neq v_{s+r+1}$  when  $s + r < m + n$ . The statistic  $U$  defined as the number of runs in the sequence  $v_1, v_2, \dots, v_{m+n}$  seems a suitable statistic for testing the null hypothesis. A difference in the distribution functions of  $X$  and  $Y$  tends to decrease  $U$ . Hence the critical region is defined by the inequality  $U < u_0$ , where  $u_0$  depends only on  $m, n$ , and the level of significance adopted. If  $m \leq n$  and  $P\{U = c\}$  is the probability that  $U = c$ , then:

$$P\{U = 2K\} = \frac{2^{(m-1)C_{k-1} \cdot n^{-1}C_{k-1}}}{m+nC_m}, \quad (K = 1, 2, \dots, m),$$

$$P\{U = 2K - 1\} = \frac{(m-1)C_{k-1} \cdot n^{-1}C_{k-2} + m^{-1}C_{k-2} \cdot n^{-1}C_{k-1}}{m+nC_m}, \quad (K = 2, 3, \dots, m+1).$$

The mean of  $U$  is:

$$\frac{2mn}{m+n} + 1.$$

The variance of  $U$  is:

$$\frac{2mn(2mn - m - n)}{(m+n)^2(m+n-1)}.$$

If  $\frac{m}{n} = \alpha$  (a positive constant) and  $m \rightarrow \infty$ , the distribution of  $U$  converges to the normal distribution.

**The Distribution of Quadratic Forms In Non-Central Normal Random Variables.** WILLIAM G. MADOW, Washington, D. C. (Presented to the Institute under a slightly different title)

Let the distribution of a sum of non-central squares of normally and independently distributed random variables which have the unit variances be called the  $\chi'^2$  distribution. It is proved that if a set of quadratic forms have a sum which is the sum of the squares of their variables, then a necessary and sufficient condition that the quadratic forms be independently distributed in  $\chi'^2$  distributions is that the rank of the sum of quadratic forms be equal to the sum of the ranks of the quadratic forms. Furthermore, the constants on which the  $\chi'^2$  distributions depend may be obtained by substituting the values about which the variables are taken for the variables themselves in the quadratic forms. Roughly speaking the theorem states that if a set of quadratic forms satisfy the conditions of the Fisher-Cochran theorem when the true means vanish, then the set of quadratic forms will be independently distributed in  $\chi'^2$  distributions when the true means do not vanish.

### Some Theoretical Aspects of the Use of Transformations in the Statistical Analysis of Replicated Experiments. W. G. COCHRAN, Iowa State College.

The device of transforming the data to a different scale before performing an analysis of variance has recently been recommended by a number of writers for replicated experiments in which the original data show a markedly skew distribution. The use of transformations to obtain an approximate analysis has been supported mainly on the grounds that in the transformed scale the true experimental error variance is approximately the same on all plots. This paper considers the relation of the method of transformations to a more exact analysis. Discussion is confined to the  $\sqrt{x}$  and  $\sin^{-1} \sqrt{x}$  transformations, which appear to receive the most frequent use in practice.

To obtain an exact analysis, it is necessary to specify (i) how the expected value on any plot is obtained from unknown parameters representing the treatment and block (or row and column) effects (ii) how the observed values on the plots vary about the expected values. If the latter variation follows the Poisson law, (a case to which the square root transformation has been considered appropriate), the equations of estimation by maximum likelihood take the form

$$(1) \quad \sum_c \left( \frac{x - m}{m} \right) \frac{\partial m}{\partial c} = 0,$$

where  $x$  is the observed and  $m$  the expected value on any plot,  $c$  is a typical unknown parameter, and the summation extends over all plots whose expectations involve  $c$ . As the number of parameters is usually large (e.g. 16 in a  $6 \times 6$  Latin square), these equations are laborious to solve; moreover, the question of obtaining small-sample tests of significance is difficult. It is shown that if a particular form can be assumed for the prediction formula in (i), namely that  $\sqrt{m}$  is a linear function of the treatment and block (or row and column) constants, the equations of estimation may be reduced to the simpler form

$$(2) \quad \sum_c 4(r' - \sqrt{m}) = 0,$$

where  $r' = \frac{1}{2} \left( \sqrt{m} + \frac{x}{\sqrt{m}} \right)$  is a function closely related to the square root of  $x$ . It follows that the statistical analysis in square roots, with some slight adjustments, coincides with the maximum likelihood solution, provided that the above form can be assumed for the prediction formula. The appropriateness of this form in practice is briefly considered and a "goodness of fit" test by  $\chi^2$  is developed. A numerical example is worked as an illustration and indicates that a good approximation is obtained by the transformation alone even with very small numbers per plot. The corresponding theory is also discussed for the inverse sine transformation, which applies where the original data are percentages or fractions whose experimental errors are derived from the binomial distribution.



In practice the type of analysis outlined above is unlikely to supplant the simple use of transformations, because it can seldom be assumed that the experimental variance is entirely of the Poisson or binomial type. The more exact analysis may, however, be useful (i) for cases in which the plot yields are very small integers or the ratios of very small integers (ii) in showing how to give proper weight to an occasional zero plot yield.

### **The Standard Errors of Geometric and Harmonic Types of Index Numbers.** By NILAN NORRIS, Hunter College.

Various statisticians have made empirical studies of the sampling errors of certain types of index numbers used in the United States and England. None of these writers has taken advantage of the tools afforded by the modern theory of estimation, including fiducial inference, as a means of arriving at direct and general expressions for estimating the standard deviations of the sampling errors of geometric and harmonic types of index numbers.

A known expression for the first approximation to the variance of a function, as given by the relation between the variance of the function and the variance of the argument, is valid for that general class of distributions of which the variance and a higher moment are finite. With the aid of this relation, there appear simple and useful forms for estimating the standard errors of geometric and harmonic types of indexes. For sufficiently large samples, these forms are valid for all of the types of distributions of price relatives, production relatives, and similar observations ordinarily encountered, provided that there are satisfied the necessary conditions for drawing sound inferences on the basis of sampling without reference to the value of the variate.

Necessary conditions for using tests of significance soundly in connection with index number problems are those of realistic and intimate acquaintance with observations, and careful attention to certain broad theoretical considerations which determine whether or not the index is suited for the purpose for which it is used.

### **A Study of R. A. Fisher's $z$ Distribution and the Related $F$ Distribution.** L. A. AROIAN, Hunter College.

The following results for the  $z$  distribution and related  $F$  distribution are investigated:

- (1) Geometric properties.
- (2) Exact values of the seminvariants and moments of  $z$ . Exact values of the first four central moments of  $F$ .
- (3) The approach to normality of both distributions as  $n_1$  and  $n_2$  become large in any manner whatever.
- (4) The Pearson types of approximating curves, the logarithmic normal approximation, the Gram-Charlier approximation, and the uses of these in finding any level of significance of  $z$  and of  $F$ .

### **A Note on the Analysis of Variance with Unequal Class Frequencies.** ABRAHAM WALD, Columbia University.

Let us consider  $p$  groups of variates and denote by  $m_j$  ( $j = 1, \dots, p$ ) the number of elements in the  $j$ -th group. Let  $x_{ij}$  be the  $i$ -th element in the  $j$ -th group. [We assume that  $x_{ij}$  is the sum of two variates  $\epsilon_{ij}$  and  $\eta_j$ , i.e.  $x_{ij} = \epsilon_{ij} + \eta_j$  where  $\epsilon_{ij}$  ( $i = 1, \dots, m_j$ ;  $j = 1, \dots, p$ ) is normally distributed with mean  $\mu$  and variance  $\sigma^2$ , and  $\eta_j$  ( $j = 1, \dots, p$ ) is normally distributed with mean  $\mu'$  and variance  $\sigma'^2$ . All the variates  $\epsilon_{ij}$  and  $\eta_j$  are supposed to be distributed independently. The intra-class correlation  $\rho$  is given by

$$\rho = \frac{\sigma'^2}{\sigma^2 + \sigma'^2}.$$

Confidence limits for  $\rho$  have been derived only in case of equal class frequencies, i.e.  $m_1 = m_2 = \dots = m_p$ . We give here the confidence limits for  $\rho$  in case of unequal class frequencies. Since  $\rho$  is a monotonic function of  $\frac{\sigma'^2}{\sigma^2}$ , it is sufficient to derive confidence limits for  $\frac{\sigma'^2}{\sigma^2}$ . Denote  $\frac{\sigma'^2}{\sigma^2}$  by  $\lambda^2$  and the arithmetic mean of the  $j$ -th group by  $\bar{x}_j$ . Let

$$w_j = \frac{m_j}{1 + m_j \lambda^2},$$

and denote by  $F_1$  and  $F_2$  the lower and upper confidence limits respectively of  $F$ , where  $F$  has the analysis of variance distribution with  $p - 1$  and  $N - p = m_1 + \dots + m_p - p$  degrees of freedom. Then the lower confidence limit  $\lambda_1^2$  of  $\lambda^2$  is given by the root of the equation in  $\lambda^2$ :

$$(1) \quad f(\lambda^2) = \frac{N - p}{p - 1} \cdot \frac{\sum_{j=1}^p \left\{ w_j \left( \bar{x}_j - \frac{\sum w_i \bar{x}_i}{\sum w_i} \right)^2 \right\}}{\sum \sum (x_{ij} - \bar{x}_j)^2} = F_2,$$

and the upper confidence limit  $\lambda_2^2$  of  $\lambda^2$  is given by the root of

$$(2) \quad f(\lambda^2) = F_1.$$

For calculating the roots of (1) and (2), we can make use of the fact that  $f(\lambda^2)$  is monotonically decreasing with increasing  $\lambda^2$ .

**An Approach to Problems Involving Disproportionate Frequencies.** BURTON D. SEELEY, Washington, D. C.

Applied mechanics offers an analysis of variance solution to problems of multiple classification involving disproportionate sub-class numbers. The quality of orthogonality may be attained in such problems by measuring the variability between classes of any one classification after centering the others. This approach, which is not limited by the number of classes or the number of classifications, treats the problem involving equal sub-class numbers as a special phase of the general analysis of variance.

**CONSTITUTION  
OF THE  
INSTITUTE OF MATHEMATICAL STATISTICS**

**ARTICLE I**

**NAME AND PURPOSE**

1. This organization shall be known as the Institute of Mathematical Statistics.
2. Its object shall be to promote the interests of mathematical statistics.

**ARTICLE II**

**MEMBERSHIP**

1. The membership of the Institute shall consist of Members, Fellows, Honorary Members, and Sustaining Members.
2. Voting members of the Institute shall be (a) the Fellows, and (b) all others who have been members for twenty-three months prior to the date of voting.

**ARTICLE III**

**OFFICERS, BOARD OF DIRECTORS, COMMITTEE ON MEMBERSHIP, AND COMMITTEE ON PUBLICATIONS**

1. The Officers of the Institute shall be a President, two Vice-Presidents, and a Secretary-Treasurer, elected for a term of one year by a majority ballot at the annual meeting of the Institute. Voting may be in person or by mail.

(a) Exception. The first group of Officers shall be elected by a majority vote of the individuals present at the organization meeting, and shall serve until December 31, 1936.

2. The Board of Directors of the Institute shall consist of the Officers and the previous President.

3. The Institute shall have a Committee on Membership composed of three Fellows. At their first meeting subsequent to the adoption of this Constitution, the Board of Directors shall elect three members as Fellows to serve as the Committee on Membership, one member of the Committee for a term of one year, another for a term of two years, and another for a term of three years. Thereafter the Board of Directors shall elect from among the Fellows one member annually at their first meeting after their election for a term of three years. The president shall designate one of the Vice-Presidents as Chairman of this Committee.

4. The Institute shall have a Committee on Publications composed of three Members or Fellows elected by the Board of Directors. The President shall designate a Vice-President as Ex Officio Chairman of this Committee.

**ARTICLE IV**

**MEETINGS**

1. A meeting for the presentation and discussion of papers, for the election of Officers, and for the transaction of other business of the Institute shall be held annually at such time as the Board of Directors may designate. Additional meetings may be called from

time to time by the Board of Directors and shall be called at any time by the President upon written request from ten Fellows. Notice of the time and place of meeting shall be given to the membership by the Secretary-Treasurer at least thirty days prior to the date set for the meeting. All meetings except executive sessions shall be open to the public. Only papers accepted by a Program Committee appointed by the President may be presented to the Institute.

2. The Board of Directors shall hold a meeting immediately after their election and again immediately before the expiration of their term. Other meetings of the Board may be held from time to time at the call of the President or any two members of the Board. Notice of each meeting of the Board, other than the two regular meetings, together with a statement of the business to be brought before the meeting, must be given to the members of the Board by the Secretary-Treasurer at least five days prior to the date set therefor. Should other business be passed upon, any member of the Board shall have the right to reopen the question at the next meeting.

3. The Committee on Membership shall hold a meeting immediately after the annual meeting of the Institute. Further meetings of the Committee may be held from time to time at the call of the Chairman or any member of the Committee provided notice of such call and the purpose of the meeting is given to the members of the Committee by the Secretary-Treasurer at least five days before the date set therefor. Should other business be passed upon, any member of the Committee shall have the right to reopen the question at the next meeting.

4. At a regularly convened meeting of the Board of Directors, three members shall constitute a quorum. At a regularly convened meeting of the Committee on Membership, two members shall constitute a quorum.

## ARTICLE V

### PUBLICATIONS

1. The *Annals of Mathematical Statistics* shall be the Official Journal for the Institute. Other publications may be originated by the Board of Directors as occasion arises.

## ARTICLE VI

### EXPULSION OR SUSPENSION

1. Except for non-payment of dues, no one shall be expelled or suspended except by action of the Board of Directors with not more than one negative vote.

## ARTICLE VII

### AMENDMENTS

1. This constitution may be amended by an affirmative two-thirds vote at any regularly convened meeting of the Institute provided notice of such proposed amendment shall have been sent to each voting member by the Secretary-Treasurer at least thirty days before the date of the meeting at which the proposal is to be acted upon. Voting may be in person or by mail.

## BY-LAWS

### ARTICLE I

#### DUTIES OF THE OFFICERS, BOARD OF DIRECTORS, COMMITTEE ON MEMBERSHIP, AND COMMITTEE ON PUBLICATIONS

1. The President, or in his absence, one of the Vice-Presidents, or in the absence of the President and both Vice-Presidents, a Fellow selected by vote of the Fellows present,

shall preside at the meetings of the Institute and of the Board of Directors. At meetings of the Institute, the presiding officer shall vote only in the case of a tie, but at meetings of the Board of Directors he may vote in all cases. At least three months before the date of the annual meeting, the President shall appoint a Nominating Committee of three members. It shall be the duty of the Nominating Committee to make nominations for Officers to be elected at the annual meeting and the Secretary-Treasurer shall notify all voting members at least thirty days before the annual meeting. Additional nominations may be submitted in writing, if signed by at least ten Fellows of the Institute, up to the time of the meeting.

2. The Secretary-Treasurer shall keep a full and accurate record of the proceedings at the meetings of the Institute and of the Board of Directors, send out calls for said meetings and, with the approval of the President and the Board, carry on the correspondence of the Institute. Subject to the direction of the Board, he shall have charge of the archives and other tangible and intangible property of the Institute. He shall send out calls for annual dues and acknowledge receipt of same; pay all bills approved by the President for expenditures authorized by the Board or the Institute; keep a detailed account of all receipts and expenditures, prepare a financial statement at the end of each year and present an abstract of the same at the annual meeting of the Institute after it has been audited by a Member or Fellow of the Institute appointed by the President as Auditor. The Auditor shall report to the President.

3. The Board of Directors shall have charge of the funds and of the affairs of the Institute, with the exception of those affairs specifically assigned to the President or to the Committee on Membership. The Board shall have authority to fill all vacancies ad interim, occurring among the Officers, Board of Directors, or in any of the Committees. The Board may appoint such other committees as may be required from time to time to carry on the affairs of the Institute.

4. The Committee on Membership shall prepare and make available through the Secretary-Treasurer an announcement indicating the qualifications requisite for the different grades of membership.

5. The Committee on Publications, under the general supervision of the Board of Directors, shall have charge of all matters connected with the publications of the Institute, and of all books, pamphlets, manuscripts and other literary or scientific material collected by the Institute. Once a year this Committee shall cause to be printed in the Official Journal the Constitution and By-Laws and a classified list of all the Members and Fellows of the Institute.

## ARTICLE II

### DUES

1. Members shall pay five dollars at the time of admission to membership and shall receive the full current volume of the Official Journal. Thereafter, Members shall pay five dollars annual dues. The annual dues of Fellows shall be five dollars. The annual dues of Sustaining Members shall be fifty dollars. Honorary Members shall be exempt from all dues.

2. Annual dues shall be payable on the first day of January of each year.

3. The annual dues of a Fellow or Member include a subscription to the Official Journal. The annual dues of a Sustaining Member include two subscriptions to the Official Journal.

4. It shall be the duty of the Secretary-Treasurer to notify by mail anyone whose dues

may be six months in arrears, and to accompany such notice by a copy of this Article. If such person fail to pay such dues within three months from the date of mailing such notice, the Secretary-Treasurer shall report the delinquent one to the Board of Directors, by whom the person's name may be stricken from the rolls and all privileges of membership withdrawn. Such person may, however, be re-instated by the Board of Directors upon payment of the arrears of dues.

### ARTICLE III

#### SALARIES

1. The Institute shall not pay a salary to any Officer, Director, or member of any committee.

### ARTICLE IV

#### AMENDMENTS

1. These By-Laws may be amended in the same manner as the Constitution or by a majority vote at any regularly convened meeting of the Institute, if the proposed amendment has been previously approved by the Board of Directors.

## DIRECTORY OF THE INSTITUTE OF MATHEMATICAL STATISTICS

(As of January 1, 1940)

Please notify the Secretary, P. R. Rider, Washington University,  
St. Louis, Mo., of any changes of addresses or errors in names, titles or  
addresses as printed.

- Anderson, Mr. Paul H. 108 E. Chalmers St., Champaign, Ill.  
Acerboni, Dr. Argentino V. Larroque 232 Banfield, Buenos Aires, Argentina.  
Alter, Prof. Dinsmore. Director of Griffith Observatory, Los Angeles, Calif.  
Arnold, Prof. Herbert E. Dept. of Math., Wesleyan Univ., Middletown, Conn.  
Aroian, Prof. Leo. Dept. of Math., Hunter College, Lexington Ave. and E. 68th St.,  
New York, N. Y.  
Bacon, Prof. H. M. Box 1144, Stanford Univ., Calif.  
Baker, Dr. G. A. Experiment Station, College of Agriculture, Univ. of Calif., Davis,  
Calif..  
Barral-Souto, Dr. José. Cordoba 1459, Buenos Aires, Argentina.  
Barrett, Mr. C. S. 3145 Maple Ave., Brookfield, Ill.  
Beall, Mr. Geoffrey. 729 Queen St., Chatham, Ontario, Canada.  
Been, Mr. Richard O. 32 Old Fisheries Bldg., U. S. Bureau of Agricultural Economics,  
Washington, D. C.  
Bennett, Prof. A. A. Dept. of Math., Brown Univ., Providence, R. I.  
Berkson, Dr. Joseph. Mayo Clinic, Rochester, Minn.  
Bernstein, Prof. Felix. Dept. of Biometry, New York Univ., New York, N. Y..  
Boley, Mr. Charles C. 305 Ceramics Bldg., Urbana, Ill.  
Bridger, Mr. Clyde A. 1801 No. Eleventh St., Boise, Idaho.  
Brooks, Mr. A. G. 5819 W. Erie St., Chicago, Ill.  
Burgess, Dr. R. W. Western Electric Co., 195 Broadway, New York, N. Y.  
Burr, Prof. Irving W. Antioch College, Yellow Springs, Ohio.  
Bushey, Prof. J. Hobart. Dept. of Math., Hunter College, Park Ave. and 68th St., New  
York, N. Y.  
Camp, Prof. B. H. Dept. of Math., Wesleyan Univ., Middletown, Conn..  
Camp, Prof. C. C. Dept. of Math., Univ. of Nebraska, Lincoln, Neb.  
Carlson, Mr. John L. 736 West St., Reno, Nev.  
Carver, Prof. H. C. Dept. of Math., Univ. of Michigan, Ann Arbor, Mich..  
Chapman, Mr. Roy A. Forest Service, U. S. Dept. of Agriculture, 1000 Masonic Temple,  
New Orleans, La.  
Clark, Prof. Andrew G. 1012 Laporte Ave., Ft. Collins, Colo.  
Cochran, Prof. W. G. Statistical Laboratory, Iowa State College, Ames, Iowa.  
Coleman, Mr. E. P. Dept. of Math., Municipal Univ. of Omaha, Omaha, Neb.  
Cox, Dr. Gerald J. Mellon Institute, Pittsburgh, Pa.  
Craig, Prof. A. T. Dept. of Math., Univ. of Iowa, Iowa City, Iowa..  
Craig, Prof. C. C. Dept. of Math., Univ. of Michigan, Ann Arbor, Mich..  
Crathorne, Prof. A. R. Dept. of Math., Univ. of Illinois, Urbana, Ill..  
Crowe, Prof. S. E. 137 University Drive, East Lansing, Mich.  
Curtiss, Dr. J. H. Belle Ayre Apts., Ithaca, N. Y.  
Dahmus, Mr. Maurice E. 203 So. Fourth, Champaign, Ill.

- Daly, Dr. Joseph F. 719 Irving St., N.E., Washington, D. C.  
 Dantzig, Mr. George B. 2609 Fulton St., Berkeley, Calif.  
 DeLury, Dr. D. B. Dept. of Math., Univ. of Toronto, Toronto 5, Canada.  
 Deming, Dr. W. E. Bureau of the Census, U. S. Dept. of Commerce, Washington, D. C..  
 Dodd, Prof. E. L. Dept. of Math., Univ. of Texas, Austin, Texas..  
 Dodge, Mr. Harold F. Bell Telephone Lab., 463 West St., New York, N. Y.  
 Doob, Dr. J. L. Dept. of Math., Univ. of Illinois, Urbana, Ill..  
 Dressel, Dr. Paul L. 6126 S. Woodlawn Ave. Chicago, Ill.  
 Dunlap, Prof. Jack W. Catharine Strong Hall, Univ. of Rochester, Rochester, N. Y.  
 Dwyer, Prof. Paul S. 407 Camden Ct., Ann Arbor, Mich.  
 Edgett, Dr. G. L. Queen's Univ., Kingston, Ontario, Canada.  
 Eisenhart, Dr. Churchill. Dept. of Math., Univ. of Wisconsin, Madison, Wis.  
 Elston, Mr. James S. Travelers Insurance Co., Hartford, Conn.  
 Elting, Mr. John P. Director of Research, Kendall Mills, Paw Creek, N. C.  
 Elveback, Miss Mary L. Princeton Surveys, Princeton, N. J.  
 Ettinger, Mr. W. J. Edison General Electric Appliance Co., 5600 W. Taylor St., Chicago, Ill.  
 Evans, Prof. H. P. North Hall, Univ. of Wisconsin, Madison, Wis.  
 Feldman, Dr. H. M. Sodan High School, St. Louis, Mo.  
 Fertig, Dr. John W. Dept. of Biostatistics, School of Hygiene and Public Health, Johns Hopkins Univ., Baltimore, Md.  
 Fischer, Dr. C. H. Dept. of Math., Wayne Univ., Detroit, Mich.  
 Fisher, Prof. Irving. 460 Prospect St., New Haven, Conn..  
 Flood, Dr. M. M. Princeton Surveys, Princeton, N. J.  
 Foster, Mr. Ronald M. 122 E. Dudley Ave., Westfield, N. J.  
 Frankel, Mr. Lester R. 1445 Otis Place, N.W., Washington, D. C.  
 Freeman, Prof. H. A. Dept. of Economics, Massachusetts Inst. of Technology, Cambridge, Mass.  
 Fry, Dr. T. C. Bell Telephone Lab., 463 West St., New York, N. Y..  
 Gavett, Prof. G. I. Dept. of Math., Univ. of Washington, Seattle, Wash.  
 Geiringer, Dr. Hilda. Bryn Mawr College, Bryn Mawr, Pa.  
 Gibson, Mr. Robert W. 503 W. California, Urbana, Ill.  
 Gill, Mr. John P. Dept. of Math., Univ. of Alabama, University, Ala.  
 Girschick, Mr. M. A. Bureau of Home Economics, Dept. of Agriculture, Washington, D. C..  
 Glover, Prof. J. W. 620 Oxford Rd., Ann Arbor, Mich..  
 Greenwood, Prof. J. A. College Station, Box 245, Durham, N. C.  
 Greville, Dr. Thos. N. E. Dept. of Math., Univ. of Michigan, Ann Arbor, Mich.  
 Grove, Prof. C. C. 143 Milburn Ave., Baldwin, L. I., N. Y.  
 Hart, Prof. W. L. Dept. of Math., Univ. of Minnesota, Minneapolis, Minn.  
 Hebley, Mr. Henry F. Product Control Manager, Pittsburgh Coal Co., P. O. Box 145, Pittsburgh, Pa.  
 Heldingsfield, Mr. Myron. Hotel Marcy, 720 West End Ave., New York, N. Y.  
 Henderson, Mr. Robert. Crown Point, Essex Co., N. Y..  
 Hendricks, Mr. Walter A. Division of Agricultural Statistics, Agricultural Marketing Service, U. S. Dept. of Agriculture, Washington, D. C.  
 Henry, Mr. M. H. Route No. 4, Box 394 M, Lansing, Mich.  
 Hoel, Dr. Paul G. Dept. of Math., Univ. of California at Los Angeles, Los Angeles, Calif.  
 Horst, Dr. Paul. Supervisor of Selection Research, Procter and Gamble, Cincinnati, Ohio.  
 Hotelling, Prof. Harold. Fayerweather Hall, Columbia Univ., New York, N. Y..  
 Huntington, Prof. E. V. 48 Highland St., Cambridge, Mass..  
 Hurwitz, Mr. William. 119 Concord Ave., Washington, D. C.  
 Ingraham, Prof. M. H. North Hall, Univ. of Wisconsin, Madison, Wis..  
 Jackson, Prof. Dunham. 119 Folwell Hall, Univ. of Minnesota, Minneapolis, Minn..



- Jacobs, Mr. Walter. 5119 Third St., N.W., Washington, D. C.  
 Janko, Dr. Jaroslav. Praha XII, Krkonosska 3, Czechoslovakia.  
 Juran, Mr. J. M. Western Electric Co., 195 Broadway, New York, N. Y.  
 Keeping, Prof. E. S. Univ. of Alberta, Edmonton, Alberta, Canada.  
 Kelley, Prof. Truman L. Lawrence Hall, Cambridge, Mass..  
 Kenney, Mr. John F. Dept. of Math., Northwestern Univ., Evanston, Ill.  
 Kiernan, Prof. Charles J. 14 Fairbanks St., Hillside, N. J.  
 Knowler, Dr. Lloyd A. Dept. of Math., Univ. of Iowa, Iowa City, Iowa.  
 Knudsen, Miss Lila F. Junior Mathematician, Food Division, U. S. Dept. of Agriculture, Washington, D. C.  
 Kohl, Miss Alma. 2309 Russell St., Berkeley, Calif.  
 Kossack, Mr. Carl F. Dept. of Math., Univ. of Oregon, Eugene, Oregon.  
 Kullback, Dr. Solomon. Office of Chief Signal Officer, U. S. Navy Department, Washington, D. C..  
 Kurtz, Dr. Albert K. 173 Cornwall St., Hartford, Conn.  
 Laderman, Mr. Jack. 3224 Bronx Blvd., Bronx, N. Y.  
 Larsen, Prof. Harold D. Dept. of Math., Univ. of New Mexico, Albuquerque, N. M.  
 Leavens, Mr. Dickson H. Cowles Commission, Univ. of Chicago, Chicago, Ill.  
 Lemme, Prof. Maurice M. Dept. of Math., Univ. of Toledo, Toledo, Ohio.  
 Lindsey, Mr. Fred D. 2051 No. Brandywine St., Arlington, Va.  
 Livers, Mr. Joe J. Dept. of Math., Montana State College, Bozeman, Mont.  
 Lorge, Dr. Irving. 525 W. 120 St., New York, N. Y.  
 Lotka, Dr. A. J. One Madison Ave., New York, N. Y..  
 Lundberg, Prof. George A. Dept. of Sociology, Bennington College, Bennington, Vt.  
 Madow, Mr. William G. Hotel Lafayette, Washington, D. C..  
 Malzberg, Dr. Benjamin. New York State Dept. of Mental Hygiene, Albany, N. Y.  
 Mansfield, Prof. Ralph. Dept. of Math., Chicago Teachers College, 6800 Stewart Ave., Chicago, Ill.  
 Mauchly, Prof. J. W. Dept. of Physics, Ursinus College, Collegeville, Pa.  
 Mayer, Mr. George F. T. 1519 Seventh St., S.E., Minneapolis, Minn.  
 Mises, Prof. R. von. Harvard Univ., Cambridge, Mass..  
 Mode, Prof. E. B. 688 Boylston St., Boston, Mass.  
 Molina, Mr. E. C. 463 West St., New York, N. Y..  
 Mudgett, Prof. Bruce D. Dept. of Economics, Univ. of Minnesota, Minneapolis, Minn.  
 MacLean, Mr. M. C. Chief of Census Analysis, Dominion Bureau of Statistics, Ottawa, Ontario, Canada.  
 McEwen, Prof. G. F. Scripps Institution, La Jolla, Calif.  
 Macphall, Dr. M. S. Dept. of Math., Acadia Univ., Wolfville, Nova Scotia, Canada.  
 McCarthy, Dr. Michael C. Dept. of Math., University College, Cork, Ireland.  
 Nesbitt, Dr. C. J. Dept. of Math., Univ. of Michigan, Ann Arbor, Mich.  
 Neyman, Prof. J. Dept. of Math., Univ. of California, Berkeley, Calif..  
 Norris, Dr. Nilan. Dept. of Economics, Hunter College, New York, N. Y.  
 Norton, Mr. K. A. 2123 Tunlaw Rd., N.W., Washington, D. C.  
 Olds, Prof. E. G. 953 La Clair Ave., Regent Square, Pittsburgh, Pa..  
 Ollivier, Dr. Arthur. Box 405, State College, Miss.  
 Olmstead, Dr. Paul S. 463 West St., New York, N. Y.  
 O'Toole, Dr. A. L. 6363 Sheridan Rd., Chicago, Ill..  
 Ozanne, Mr. Paul. Dept. of Agronomy, College of Agriculture, Univ. of Wisconsin, Madison, Wis.  
 Ore, Prof. Oystein. Dept. of Math., Yale Univ., New Haven, Conn.  
 Palley, Mr. A. 286 Commonwealth, Buffalo, N. Y.  
 Parente, Mr. A. R. 126 Church St., Hamden, Conn.  
 Perryman, Mr. James H. 12 Richmond Ave., Ilfracombe, Devon, England.

- Petrie, Prof. George W. South Dakota School of Mines, Rapid City, S. Dak.  
 Pierce, Prof. J. A. Dept. of Math., Atlanta Univ., Atlanta, Ga.  
 Pixley, Prof. H. H. Dept. of Math., Wayne Univ., Detroit, Mich.  
 Pollard, Prof. H. S. Dept. of Math., Miami Univ., Oxford, Ohio.  
 Reed, Prof. L. J. 615 No. Wolfe St., Baltimore, Md.  
 Rider, Prof. P. R. Dept. of Math., Washington Univ., St. Louis, Mo..  
 Richardson, Prof. C. H. Bucknell Univ., Lewisburg, Pa.  
 Regan, Prof. Francis. Dept. of Math., St. Louis Univ., Grand and Pine Blvd., St. Louis, Mo.  
 Rietz, Prof. H. L. Dept. of Math., Univ. of Iowa, Iowa City, Iowa..  
 Romig, Mr. H. G. Bell Telephone Lab., 463 West St., New York, N. Y.  
 Roos, Dr. Charles F. Inst. of Applied Econometrics, 420 Lexington Ave., New York, N. Y..  
 Rulon, Prof. P. J. 13 Kirkland St., Cambridge, Mass.  
 Scarborough, Prof. J. B. P. O. Box 86, Annapolis, Md.  
 Schwartz, Mr. Herman M. College of Arts and Sciences, Duquesne Univ., Pittsburgh, Pa.  
 Shaw, Mr. Lawrence W. Fayerweather Hall, Columbia Univ., New York, N. Y.  
 Shewhart, Dr. W. A. 158 Lake Drive, Mountain Lakes, N. J..  
 Shohat, Prof. J. Dept. of Math., Univ. of Pennsylvania, Philadelphia, Pa..  
 Simon, Capt. Leslie E. Aberdeen Proving Ground, Aberdeen, Md.  
 Singleton, Mr. Robert R. 20 Nassau St., Princeton, N. J.  
 Spiegelman, Mr. Mortimer. 325 W. 86th St., New York, N. Y.  
 Stephan, Mr. F. F. 1626 K St., N. W., Washington, D. C.  
 Stouffer, Prof. S. A. Dept. of Sociology, Univ. of Chicago, Chicago, Ill.  
 Swanson, Mr. A. G. General Motors Inst., Flint, Mich.  
 Tai, Mr. Shih-Kuang. Tsing Hua Univ., Kun Ming, Yun Nan, China.  
 Thompson, Dr. William R. 27 Oakwood St., Albany, N. Y.  
 Thurstone, Prof. L. L. Dept. of Psychology, Univ. of Chicago, Chicago, Ill.  
 Toops, Prof. Herbert A. 458 W. 8th Ave., Columbus, Ohio.  
 Torrey, Miss Mary N. 463 West St., New York, N. Y.  
 Treloar, Prof. Alan E. Dept. of Biometry, Univ. of Minnesota, Minneapolis, Minn.  
 Trimble, Mrs. Anne. Apt. 612, 6104 Woodlawn Ave., Chicago, Ill.  
 Trubridge, Dr. G. F. P. 98 Kingsway, Petts Wood, Kent, England.  
 Vatnsdal, Mr. J. R. 304 Columbia St., Pullman, Wash.  
 Vickery, Dr. C. W. 304 E. 5th St., Austin, Texas.  
 Walte, Dr. Warren C. Division of Agricultural Economics, Univ. Farm, St. Paul, Minn.  
 Walker, Prof. Helen M. Teachers College, Columbia Univ., New York, N. Y.  
 Wald, Dr. A. Fayerweather Hall, Columbia Univ., New York, N. Y..  
 Wareham, Mr. Ralph E. 118 Elmer Ave., Schenectady, N. Y.  
 Wei, Dzung-shu. International House, 19 Evans St., Iowa City, Iowa.  
 Weida, Prof. F. M. Dept. of Math., George Washington Univ., Washington, D. C..  
 Welker, Mr. E. L. 160 Mathematics Bldg., Univ. of Illinois, Urbana, Ill.  
 Wescott, Mr. Mason E. 1936 Greenwood Ave., Wilmette, Ill.  
 White, Prof. A. E. Dept. of Math., Kansas State College, Manhattan, Kan.  
 Wilder, Dr. Marian. 115 Bedford S. E., Minneapolis, Minn.  
 Wilks, Prof. S. S. Fine Hall, Princeton Univ., Princeton, N. J..  
 Wilson, Dr. Elizabeth. One Waterman St., Cambridge, Mass.  
 Wilson, Mr. W. P. P. O. Box 1939, University, Ala.  
 Wolfenden, Mr. Hugh H. 182 Rosedale Heights Drive, Toronto 5, Canada.  
 Wright, Prof. Sewall. Dept. of Zoology, Univ. of Chicago, Chicago, Ill..  
 Wyckoff, Mr. J. F. Dept. of Math., Trinity College, Hartford, Conn.  
 Zoch, Mr. Richmond T. Cosmos Club, 1520 H. St., Washington, D. C.  
 Zubin, Dr. Joseph. 722 W. 168th St., New York, N. Y.



# LIMITING DISTRIBUTIONS OF QUADRATIC AND BILINEAR FORMS<sup>1,2</sup>

BY WILLIAM G. MADOW

**1. Introduction.** In a previous paper [15], several generalizations of the theorem of Fisher, [6, p. 97] and Cochran, [2, p. 178] on the joint distribution of quadratic forms in normally and independently distributed random variables were derived. The chief purpose of this paper is a demonstration that the Fisher-Cochran theorem and its generalizations are valid in the limit under conditions completely analogous to those under which the Laplace-Liapounoff theorem holds. Applications to the analysis of variance, periodogram analysis and multivariate analysis are discussed.

Our general procedure will be to find algebraic conditions on the matrices of quadratic and bilinear forms which enable us to assert that the limiting distributions of these forms are those which they would have had if the variables, the squares or products of which appear in their canonical forms, had been normally and independently distributed.<sup>3</sup> One thing which makes this possible is the fact that many frequently used quadratic and bilinear forms have the same rank no matter what may be the number of variables of which they are functions. For example, the rank of the square of the arithmetic mean,  $\bar{x}_n$ , where

$$\bar{x}_n = \frac{1}{n}(x_1 + \dots + x_n),$$

is one for all values of  $n$ . In this case the quadratic form,

$$\frac{1}{n^2} \sum_{\mu, \nu=1}^n x_{\mu} x_{\nu},$$

is a function of the  $n$  variables  $x_1, x_2, \dots, x_n$ .

In paragraph 2 we state the vector form of the Laplace-Liapounoff theorem and several corollaries. The joint limiting distributions of quadratic and bilinear forms are derived in paragraph 3. The final paragraph is devoted to a statement of a few applications of the theorems.

<sup>1</sup> Much of this research was done under a grant-in-aid from the Carnegie Corporation of New York.

<sup>2</sup> The material contained in this paper was presented in part to the American Statistical Association, December 28, 1937, and in part to the Institute of Mathematical Statistics, December 27, 1938.

<sup>3</sup> We shall be chiefly concerned with conditions under which the limiting distributions are not themselves normal. If the limiting distributions are normal, then generally under the conditions we state, the Laplace-Liapounoff theorem will have been directly applicable.

**2. The Laplace-Liapounoff theorem.**<sup>4</sup> We shall first state some definitions and terminology which will be used throughout the paper.

If used as subscripts or superscripts, or as indices of summation or multiplication, the letters  $i, j$  will take on all integral values from 1 through  $p$ , the letters  $\mu, \nu$  will take on all integral values from 1 through  $n$ , the letters  $\gamma, \delta$  will take on all integral values from 1 through  $m$ , the letter  $\alpha$  will take on all integral values from 1 through  $k$ , and the letter  $\beta$  will take on all integral values from 1 through  $k-1$ , unless explicit statement to the contrary is made.

The totality of all sets of  $\nu$  real numbers will be denoted by  $R^\nu$ . Thus  $R^\nu$  is the combinatory product of the spaces  $R^1, R^1, \dots, R^1$ , ( $\nu$  times).

If  $x_1, \dots, x_n$  are random variables, and if  $\Delta$  is a proposition concerning  $x_1, \dots, x_n$ , then by  $P\{\Delta\}$  we shall mean "the probability that  $\Delta$ ." The distribution function of the random variables  $x_1, \dots, x_n$  will be denoted by  $F(x_1, \dots, x_n)$ , i.e.

$$F(x_1^0, \dots, x_n^0) = P\{x_1 < x_1^0, \dots, x_n < x_n^0\}$$

for all sets of  $n$  real numbers. Thus  $F$  will have an operational meaning in this paper.

If  $\Delta(x_1, \dots, x_n)$  is a function of  $x_1, \dots, x_n$  defined on  $R^n$  and measurable<sup>5</sup> with respect to  $F(x_1, \dots, x_n)$ , then  $E\{\Delta(x_1, \dots, x_n)\}$  will be defined by the equation,

$$E\{\Delta(x_1, \dots, x_n)\} = \int_{R^n} \Delta(x_1, \dots, x_n) dF(x_1, \dots, x_n),$$

where the integral is a Lebesgue-Stieltjes or Radon integral. Hence  $|\Delta(x_1, \dots, x_n)|$  is assumed to be integrable with respect to  $F(x_1, \dots, x_n)$ .

If  $\Omega(y_1, \dots, y_p)$  is a single valued measurable function of  $y_1, \dots, y_p$  on  $R^p$ , and if  $y_i$  is a real single valued Borel measurable<sup>6</sup> function of  $x_1, \dots, x_n$  on  $R^n$ , then upon substituting for  $y_1, \dots, y_p$  it is seen that  $\Omega(y_1, \dots, y_p)$

<sup>4</sup> Although the theorems will be stated in terms of probability distributions, Borel measurability, and Lebesgue-Stieltjes integrability, it may simplify the reading if the words "probability distributions" are replaced by probability densities or statistical distributions, "Borel measurability" are replaced by continuity, and "Lebesgue-Stieltjes integrability" are replaced by Riemann integrability.

<sup>5</sup> A function  $\Delta(x_1, \dots, x_n)$  defined on  $R^n$  is said to be measurable with respect to a distribution function  $F(x_1, \dots, x_n)$  if the set  $E(t)$  of all  $x_1, \dots, x_n$  such that  $\Delta(x_1, \dots, x_n) < t$  is such that  $\int_{E(t)} dF(x_1, \dots, x_n)$  is defined for all  $t$ .

<sup>6</sup> All subsets of  $R^n$  which may be formed from the totality of intervals of  $R^n$  by repeated summations or multiplications of not more than a denumerable number of intervals of  $R^n$ , and  $R^n$  itself, constitute the totality of Borel sets of  $R^n$ . The function  $y(x_1, \dots, x_n)$ , defined on  $R^n$ , is a Borel measurable function of  $x_1, \dots, x_n$  on  $R^n$  if the set of values of  $x_1, \dots, x_n$  such that  $y(x_1, \dots, x_n) < t$  is a Borel set for all  $t$ . The class of continuous functions is contained in the class of Borel measurable functions. For further details, see [3, chs. 1, 2], [11, ch. 3] and [17, chs. 1, 2, 3].

is a single-valued measurable function,  $\Delta(x_1, \dots, x_n)$  of  $x_1, \dots, x_n$  on  $E^n$ . If  $x_1, \dots, x_n$  are random variables, then  $y_1, \dots, y_p$  are random variables, and<sup>7</sup>

$$(2.1) \quad E\{\Omega(y_1, \dots, y_p)\} = E\{\Delta(x_1, \dots, x_n)\}.$$

We shall call  $E(x_i)$  the mean value of  $x_i$ ,  $\sigma_{ij}$  the covariance of  $x_i$  and  $x_j$ , and  $\sigma_{ii}$  or  $\sigma_i^2$  the variance of  $x_i$ , where  $\sigma_{ij} = E\{(x_i - Ex_i)(x_j - Ex_j)\}$ .

The Laplace-Liapounoff, or Central Limit theorem states conditions under which linear functions of random variables have a normal limiting distribution. The general characteristic of the proofs of the theorem is that conditions are placed on the random variables so that they may virtually be assumed to be bounded. The Lindeberg<sup>8</sup> condition, which we shall use, is perhaps the least restrictive of all the conditions which require finite means and variances.

The Lindeberg condition<sup>9</sup>,  $\mathcal{L}_p$ : A set of random variables  $x_{i\nu n}$  will be said to satisfy the Lindeberg condition  $\mathcal{L}_p$  if there exists, for any preassigned positive real numbers  $\delta$  and  $\epsilon$ , a positive integer  $n_0$  such that if  $n > n_0$ , then

$$\sum_i \int_{|z_{i\nu n}| > \epsilon} z_{i\nu n}^2 dF(x_{1\nu n}, \dots, x_{p\nu n}) < \delta,$$

where

$$z_{i\nu n}^2 = x_{1\nu n}^2 + x_{2\nu n}^2 + \dots + x_{p\nu n}^2$$

and

$$\sigma_{i1n}^2 + \sigma_{i2n}^2 + \dots + \sigma_{in}^2 = 1.$$

If

$$x_{i\nu n} = \frac{x_{i\nu}}{s_{in}} \quad \text{where} \quad s_{in}^2 = \sigma_{i1}^2 + \dots + \sigma_{in}^2,$$

and the  $x_{i\nu n}$  satisfy  $\mathcal{L}_p$  then we shall say that the  $x_{i\nu}$  satisfy  $\mathcal{L}_p$ .

Suppose that the random variables  $y_{11}, \dots, y_{pmp}$  have a normal multivariate distribution with zero means and with covariance parameters  $\sigma_{i\gamma j\delta}$  where

$$\sigma_{i\gamma j\delta} = E(y_{i\gamma} y_{j\delta}), \gamma = 1, \dots, m_i; \delta = 1, \dots, m_j,$$

and denote the distribution function of  $y_{11}, \dots, y_{pmp}$  by  $N(y)$ . Then we may state the Laplace-Liapounoff theorem as:

<sup>7</sup> It is noted that  $\Omega(y_1, \dots, y_p)$  is integrated with respect to  $F(y_1, \dots, y_p)$  and  $\Delta(x_1, \dots, x_n)$  is integrated with respect to  $F(x_1, \dots, x_n)$ .

<sup>8</sup> See Cramer [3, pp. 57, 60, 114], and the references there given.

<sup>9</sup> It is not difficult to show that the Lindeberg condition will be satisfied if moments of order greater than two exist, [3, p. 60], or if the conditions stated by Levy [13, p. 207] and [14, p. 106] are satisfied.

**THEOREM I.** Suppose that, for each value of  $n$ , the random variables  $x_{i\gamma n}$ , which are independent for different values of  $\gamma$ , have zero means and covariance parameters  $\sigma_{i\gamma j\delta n}$ , where

$$\sigma_{i\gamma j\delta n} = E(x_{i\gamma n} x_{j\delta n}).$$

Denote by  $d'_n$  the maximum of the variances  $\sigma_{i\gamma i\gamma n}$ . If the functions  $y_{i\gamma n}$  are defined by the equations

$$y_{i\gamma n} = \sum_{\gamma} x_{i\gamma n},$$

it follows that

$$\sigma_{i\gamma j\delta n} = E(y_{i\gamma n} y_{j\delta n}) = \sum_{\gamma} \sigma_{i\gamma j\delta n}.$$

If  $\lim_{n \rightarrow \infty} \sigma_{i\gamma j\delta n} = \sigma_{i\gamma j\delta}$  and if  $\lim_{n \rightarrow \infty} d'_n = 0$ , then a necessary and sufficient condition that as  $n \rightarrow \infty$ , the limiting distribution<sup>10</sup> of  $y_{11n}, \dots, y_{pm_p n}$  be  $N(y)$  is that the condition  $\mathfrak{L}_{p, m_p}$  be satisfied.

The proof of this theorem is omitted. It may readily be developed from the proofs of Cramer, [3, pp. 57, 113].

Before stating certain corollaries which are of interest, some additional definitions are necessary.

Let  $C_n, C_{n+1}, \dots$  be a sequence of  $m$  rowed real matrices

$$C_n = ||c_{\gamma n}||, \quad n = m, m+1, \dots,$$

and let the greatest of the absolute values of the elements of  $C_n$  be denoted by  $d_n$ . The inner product of any two rows of  $C_n$  will be denoted by  $\rho_{\gamma\delta n}$ , i.e.

$$\rho_{\gamma\delta n} = \sum_{\gamma} c_{\gamma n} c_{\delta n}.$$

Let  $X_1, X_2, \dots$  be a sequence of random vectors of  $p$  components defined on  $R^p$ , and let the components of  $X_\mu$  be denoted by  $x_{1\mu}, \dots, x_{p\mu}$ . Let the components of the chance matrix  $Y_n = ||y_{i\gamma n}||$  which has  $p$  rows and  $m$  columns, be defined by the equations

$$(2.2) \quad y_{i\gamma n} = \sum_{\gamma} c_{\gamma n} x_{i\gamma}$$

for each value of  $n$ , ( $n = m, \dots; m \geq p$ ).

<sup>10</sup> The distribution functions  $F(X_n)$  will be said to converge to the distribution function  $F(X)$  if and only if

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} dF(X_n) = F(X)$$

for every  $X$  at which  $F(X)$  is continuous. If  $F(X)$  is continuous throughout  $R^n$ , then the convergence is uniform.

Suppose that

$$(2.3) \quad E(x_{i\mu}) = 0$$

and

$$(2.4) \quad E(x_{i\mu}x_{j\nu}) = \sigma_{ij}\delta_{\mu\nu},$$

where  $\delta_{\mu\nu} = 1$  if  $\mu = \nu$  and  $\delta_{\mu\nu} = 0$  if  $\mu \neq \nu$ . (There should be no confusion of this use of the letter  $\delta$  with its use as an index.) It is easy to see that if the  $c_{\gamma\mu}$  are real numbers, then

$$E(y_{i\gamma\mu}) = 0$$

and

$$E(y_{i\gamma\mu}y_{j\delta\nu}) = \sigma_{ij}\rho_{\gamma\delta}\delta_{\mu\nu}.$$

Let the determinant of the positive definite symmetric matrix,  $(\sigma) = \|\sigma_{ij}\|$  be denoted by  $\sigma$ . Let the inverse matrix of  $(\sigma)$  be denoted by  $(\sigma)^{-1} = \|\sigma^{ij}\|$  where  $\sigma^{ij}$  is the cofactor of  $\sigma_{ij}$  in  $(\sigma)$  divided by  $\sigma$ . The determinant of  $(\sigma)^{-1}$  is  $\sigma^{-1}$ .

By  $N_d(x_1, \dots, x_p; (\sigma))$  we shall mean the normal probability density with zero means and covariance parameters  $\sigma_{ij}$ , i.e.,

$$N_d(x_1, \dots, x_p; (\sigma)) = (2\pi\sigma)^{-1} \exp \left[ -\frac{1}{2} \sum_{i,j} \sigma^{ij} x_i x_j \right], \quad (-\infty < x_i < \infty),$$

where  $(\sigma)$  is a positive definite matrix. If the random variables  $x_1, \dots, x_p$  have probability density  $N_d(X; (\sigma)) \equiv N_d(x_1, \dots, x_p; (\sigma))$ , where  $X$  is a vector, then we shall say that  $X$  has a distribution function  $N(X; (\sigma))$ , i.e.

$$\frac{\partial^p}{\partial x_1 \dots \partial x_p} N(X; (\sigma)) = N_d(X; (\sigma))$$

or

$$\int_{-\infty}^{x_p} \dots \int_{-\infty}^{x_1} N_d(t_1, \dots, t_p; (\sigma)) dt_1 \dots dt_p = N(X; (\sigma)).$$

Inasmuch as certain hypotheses will be used on several occasions in this paper, they are stated here.

If  $x_1, x_2, \dots$  are independently distributed, if (2.3) and (2.4) hold and if the  $x$ 's satisfy the condition  $\mathcal{L}_p$  then we shall say that  $\mathcal{K}_p$  is true.

If  $C_n$  is such that, for all  $n$ , the equations  $\rho_{\gamma i n} = \delta_{\gamma i}$  are true, we shall say that  $\mathcal{C}$  is true.

The following corollary is useful in deriving limiting distributions in the analysis of variance.

**COROLLARY I.** *Let  $\mathcal{K}_p$  and  $\mathcal{C}$  be true. Then a sufficient condition that*

$$\lim_{n \rightarrow \infty} F(Y_n) = \prod_{\gamma} N(y_{1\gamma}, \dots, y_{p\gamma}; (\sigma))$$

is  $\lim d_n = 0$ .



The proof is based on the fact that the  $x_{i, r, n}$  of Theorem I are given by  $c_{r, r, n} x_{i, r}$ . The details are omitted.

The  $pm$  rowed square matrix,  $(\tau) = \|\tau_{rs}\|$  is defined as follows: If  $r \leq m$ ,  $s \leq m$ ; then  $\tau_{rs} = \sigma_{11} \rho_{rs}$ ; and if  $km < r \leq (k+1)m$ ,  $lm < s \leq (l+1)m$ ,  $l, k = 0, \dots, p-1$ , then  $\tau_{rs} = \sigma_{k+1, l+1} \rho_{r-km, s-lm}$ . The inverse matrix of  $(\tau)$ , and the determinants of  $(\tau)$  and  $(\tau)^{-1}$  are defined as are  $(\sigma)^{-1}$ ,  $\sigma$  and  $\sigma^{-1}$ .

COROLLARY II. Let  $\mathcal{H}_p$  be true, and let

$$\lim_{n \rightarrow \infty} \rho_{r, n} = \rho_{rs}, \quad \rho_{rs} = 1.$$

Then, if  $\lim_{n \rightarrow \infty} d_n = 0$ , it follows that

$$\lim_{n \rightarrow \infty} F(Y_n) = F(Y),$$

where  $F(Y)$  is the distribution function determined by the probability density

$$(2\pi)^{-\frac{pm}{2}} \tau^{-1} \exp \left[ -\frac{1}{2} \sum_{r,s=1}^{pm} \tau^{rs} y_{k+1, r-km} y_{l+1, s-lm} \right]$$

where, if  $r \leq m$ ,  $s \leq m$ , then  $k = 0$ ,  $l = 0$ ; if  $r \leq m$ ,  $m < s \leq 2m$ , then  $k = 0$ ,  $l = 1$ ; and so on.

The proof is omitted.

If  $Z_1, \dots, Z_t$  are random variables, then  $F(X_1, \dots, X_k | Z_1, \dots, Z_t)$  is the distribution function of the random vectors  $X_1, \dots, X_k$  for fixed values of  $Z_1, \dots, Z_t$ , i.e. for any fixed values of  $Z_1, \dots, Z_t$ ,

$$P\{X_1 < X_1, \dots, X_k < X_k\} = F(X_1, \dots, X_k | Z_1, \dots, Z_t).$$

We shall now assume that the elements  $c_{r, r, n}$  of the matrix  $C_n$  are Borel measurable functions of a set of random variables<sup>11</sup>  $Z_1, \dots, Z_{t_n}$ . Then the matrix  $C_n$  may be called a random matrix defined on a space  $W_n$  which is the combinatory product of the spaces on which  $Z_1, \dots, Z_{t_n}$  are defined. If, for each value of  $n$ , and for all  $X^n$  and  $Z^n$ , the equation

$$(2.5) \quad F(X^n, Z^n) = F(Z^n) \cdot \prod_i F(X_i | Z_i)$$

is satisfied, then we shall say that  $\mathcal{J}$  is true. It is obvious that sufficient conditions for the truth of  $\mathcal{J}$  are

$$F(X^n, Z^n) = F(Z^n) \cdot \prod_i F(X_i),$$

or, if  $t_n \geq n$

$$F(X^n, Z^n) = F(Z_{n+1}, \dots, Z_{t_n}) \cdot \prod_i F(X_i, Z_i)$$

<sup>11</sup> The symbol  $X^n$  will stand for the set of variables  $X_1, \dots, X_n$ , and the symbol  $Z^n$  will stand for the set of variables  $Z_1, \dots, Z_{t_n}$ .

or, if  $l_n \leq n$

$$F(X^n, Z^n) = \prod_{r=1}^{l_n} F(X_r, Z_r) \cdot \prod_{r=l_n+1}^n F(X_r).$$

Inasmuch as we shall often use Fubini's theorem, it is now stated here.<sup>12</sup>

**THEOREM II.** *Let the distribution function of  $X^n, Z^n$  be  $F(X^n, Z^n)$ , let the distribution function of  $X^n$  for fixed values of  $Z^n$  be  $F(X^n | Z^n)$ , and let the distribution function of  $Z^n$  be  $F(Z^n)$ . Then if  $\Delta(X^n, Z^n)$  is measurable with respect to  $F(X^n, Z^n)$  and if*

$$\int_{R^{pn} \times W_n} |\Delta(X^n, Z^n)| dF(X^n, Z^n) < \infty,$$

it follows that

$$\int_{R^{pn}} |\Delta(X^n, Z^n)| dF(X^n | Z^n) < \infty$$

for almost all<sup>13</sup> sets of values of  $Z^n$  and

$$\int_{R^{pn} \times W_n} \Delta(X^n, Z^n) dF(X^n, Z^n) = \int_{W_n} \left[ \int_{R^{pn}} \Delta(X^n, Z^n) dF(X^n | Z^n) \right] dF(Z^n).$$

In Corollary I an important condition was that the maximum of the absolute values of the elements of  $C_n$  should approach zero as  $n$  increased. In order to obtain a similar condition when the elements of  $C_n$  are random variables, we shall define the function  $d(C_n)$  as follows: For each value of  $Z^n$  let  $d(C_n)$  be the maximum of the absolute values of the elements of  $C_n$ . We shall denote  $d(C_n)$  by  $d_n$ . If the elements of  $C_n$  are Borel measurable functions then  $d_n$  is a Borel measurable function of  $Z^n$ . Hence  $d_n$  is a random variable defined on  $W_n$ .

A sequence of random variables  $d_1, d_2, \dots$  is said to converge in probability to zero if, given  $\epsilon > 0$ , then

$$\lim_{n \rightarrow \infty} P\{|d_n| > \epsilon\} = 0.$$

If the sequence of functions  $d_p, d_{p+1}, \dots$  converges in probability to zero we shall say that  $\mathfrak{Z}$  is true.

If  $\mathcal{J}$  is true, and if, for almost all values of  $Z^n$  we have

$$(2.6) \quad \int_{R^p} x_i dF(X_r, Z^n) = 0,$$

$$(2.7) \quad \int_{R^p} x_i x_j dF(X_r, Z^n) = \sigma_{ij},$$

<sup>12</sup> Proofs of Fubini's theorem with the required amount of generality will be found in [5, p. 101] and [14, p. 73].

<sup>13</sup> A proposition concerning random variables is said to be true for almost all values of the variables, if it is true for all values of the variables, except perhaps for a set of probability zero with respect to the distribution function of the random variables.

and the condition  $\mathfrak{L}_p$  is satisfied with respect to the  $X$  and the distribution functions  $F(X, Z^n)$  then we shall say that  $\mathfrak{K}_p^0$  is true.

If

$$(2.8) \quad \sum_i \int_{R^p \times W_n} c_{\gamma n} c_{i \gamma n} x_{i\gamma} x_{j\gamma} dF(X, Z^n) = \sigma_{ij} \delta_{\gamma i},$$

then we shall say that  $\mathcal{C}^0$  is true. It is noted that if  $\mathcal{G}$  and (2.7) are true, then  $\mathcal{C}^0$  is true if  $\mathcal{C}$  is true for almost all sets of fixed values of  $Z^n$ .

COROLLARY III. Let  $\mathcal{C}^0$ ,  $\mathcal{G}$  and  $\mathfrak{K}_p^0$  be true. Then, if  $\mathfrak{Z}$  is true, it follows that

$$\lim_{n \rightarrow \infty} F(Y_n) = \prod_{\gamma} N(y_{1\gamma}, \dots, y_{p\gamma}; (\sigma)).$$

PROOF. It is necessary to show that the condition  $\mathfrak{L}_{pm}$  is satisfied by the variables  $c_{\gamma n} x_{i\gamma}$  if the condition  $\mathfrak{L}_p$  is satisfied by the variables  $x_{i\gamma}$  and that the condition  $\mathfrak{Z}$  implies that  $\lim_{n \rightarrow \infty} d_n = 0$  when the  $x_{i\gamma n}$  of Theorem I are set equal to the  $c_{\gamma n} x_{i\gamma}$  of Corollary III.

If we let  $\Delta_{\gamma n}^2 = \sum_{i,j} (c_{\gamma n} x_{i\gamma})^2$ ,  $\Delta_n^2 = \sum_{\gamma} \Delta_{\gamma n}^2$  and let  $s_n^2 = E\{\Delta_n^2\}$ , then, by (2.8), it is true that

$$s_n^2 = \sum_{i,j} \sigma_{ii} = m \sum_i \sigma_{ii}.$$

From  $\mathfrak{K}_p^0$  and the fact that for sufficiently large  $n$ ,  $|d_n^2(Z^n)| < 1$  for almost all  $Z^n$  we have for any preassigned  $\epsilon$  and  $\delta$ ,

$$\frac{1}{s_n^2} \int_{\Delta_n > \epsilon s_n} \Delta_n^2 dF(X^n, Z^n) \leq \frac{1}{s_n^2} \sum_{\gamma} \int_{\Delta_n > \epsilon s_n} m d_n^2(Z^n) \sum_i x_{i\gamma} dF(X, Z^n) < \delta$$

for sufficiently large  $n$ , since the set of  $x$ 's and  $Z^n$  for which  $\sum_{i,\gamma} x_{i\gamma}^2 > \epsilon s_n$  contains almost all the  $x$ 's and  $Z^n$  for which  $\Delta_n > \epsilon s_n$ . Hence, the condition  $\mathfrak{L}_{pm}$  is satisfied by the random variables  $c_{\gamma n} x_{i\gamma}$  with respect to the distribution functions  $F(X, Z^n)$ .

We now show that

$$\lim_{n \rightarrow \infty} [\max E\{(c_{\gamma n} x_{i\gamma})^2\}] = 0.$$

It is clearly true that

$$E\{(c_{\gamma n} x_{i\gamma})^2\} \leq \int_{R^p \times W_n} d_n^2 x_{i\gamma}^2 dF(X, Z^n).$$

Since  $d_n$  converges in probability to zero, and since  $d_n^2 \leq 1$  for almost all  $Z$ , we can, for any  $\epsilon > 0$ , take  $n_0$  so large that if  $n > n_0$ , then  $P\{d_n^2 > \frac{1}{2}\epsilon\} < \frac{1}{2}\epsilon$ . If  $E$  is the set on which  $d_n^2 > \frac{1}{2}\epsilon$ , we then have for all  $n > n_0$ , using (2.7),

$$\begin{aligned} E\{(c_{\gamma n} x_{i\gamma})^2\} &\leq \int_E \left[ \int_{R^p} x_{i\gamma}^2 dF(X, |Z^n) \right] dF(Z^n) \\ &\quad + \frac{\epsilon}{2} \int_{W_n} \left[ \int_{R^p} x_{i\gamma}^2 dF(X, |Z^n) \right] dF(Z^n) \leq \epsilon \sigma_{ii} \end{aligned}$$

and this inequality is also satisfied for all  $n > n_0$ .

The following discussion is useful in obtaining the limiting distributions of statistics which occur in multivariate statistical analysis.

The letter  $f$  will assume all integral values from 1 through  $s$ , the letters  $\mu, \nu$  will assume all integral values from 1 through  $n_f$ , and the letters  $\gamma, \delta$  will assume all integral values from 1 through  $m_f$ , for any  $f$ .

Let  $X_1^f, \dots$  be, for any fixed  $f$ , a sequence of random vectors of  $p_f$  components defined on  $R^{p_f}$ , and let the set of random variables  $X_1^f, \dots$  be independently distributed for any fixed  $f$ .

If, for each set of values of  $n_1, \dots, n_s$ , ( $t_n$  is a function of  $n_1, \dots, n_s$ ),

$$F(X_1^f, \dots, X_{t_n}^f, Z_1, \dots, Z_{t_n}) = \prod_f \prod_{\nu} F(X_{\nu}^f | Z_1, \dots, Z_{t_n}) \cdot F(Z_1, \dots, Z_{t_n}),$$

we shall say that  $\mathcal{A}_n$  is true.

Let, for any fixed value of  $f$ , the matrix<sup>14</sup>  $C_n^f = ||c'_{\gamma\mu n}||$  where the  $c'_{\gamma\mu n}$  are Borel measurable functions of  $X_{\mu}^f$ , ( $k < f$ ), and<sup>15</sup>  $Z^n$ , have the same properties as  $C_n$ , and let  $d(C_n^f)$  be the same function of  $C_n^f$  that  $d(C_n)$  is of  $C_n$ . We shall denote  $d(C_n^f)$  by  $d_n^f$ .

Let

$$y'_{i\gamma n} = \sum_{\mu} c'_{\gamma\mu n} x'_{i\mu}$$

and let  $Y_n^f = ||y'_{i\gamma n}||$ .

For fixed  $f$ , the  $p_f$  rowed square matrix  $(\sigma_f)$ , its inverse, and so on are defined as were the same functions of the  $\sigma_{ij}$  earlier in this paragraph but with  $\sigma_{ij}$  replacing  $\sigma_{ij}$ , where

$$E\{x'_{i\gamma}\} = 0$$

and

$$E\{x'_{i\gamma} x'_{j\delta}\} = \sigma_{ijf}.$$

If  $\mathcal{A}_n$  is true, and if for almost all values of  $Z^n$  we have

$$(2.9) \quad \int_{R^{p_f}} x'_{i\gamma} dF(X_{\nu}^f, Z^n) = 0,$$

$$(2.10) \quad \int_{R^{p_f}} x'_{i\gamma} x'_{j\delta} dF(X_{\nu}^f, Z^n) = \sigma_{ijf},$$

and the condition  $\mathcal{L}_{p_f}$  is satisfied with respect to the  $X_{\nu}^f$  and the distribution functions  $F(X_{\nu}^f, Z^n)$  then we shall say that  $\mathcal{H}_{p_f}^f$  is true.

If

$$(2.11) \quad \sum_{\gamma} \int c'_{\gamma\mu n} c'_{\delta\nu n} x'_{i\mu} x'_{j\nu} dF(X_{\nu}^f, Z^n) = \sigma_{ijf} \delta_{\gamma\delta},$$

<sup>14</sup> The superscripts  $f$  and  $k$  will not indicate multiplication but will only be indices.

<sup>15</sup> See footnote 11.

then we shall say that  $\mathcal{C}'$  is true. It is noted that if  $\mathcal{I}_f$  and (2.10) are true then  $\mathcal{C}'$  is true if  $\mathcal{C}$  is true for almost all sets of fixed values of  $X_1^1, \dots, X_n^{j-1}, Z^n$ .

If  $d_n'$  converges in probability to zero as  $n$  increases we shall say that  $\mathcal{Z}_f$  is true.

**COROLLARY IV.** *Let  $\mathcal{C}^*$ ,  $\mathcal{I}_*$  and  $\mathcal{K}_{p_1}^1, \dots, \mathcal{K}_{p_s}^s$  be true. Then, if  $\mathcal{Z}_1, \dots, \mathcal{Z}_s$  are true, it follows that*

$$\lim_{n_1, \dots, n_s \rightarrow \infty} F(Y_{n_1}^1, \dots, Y_{n_s}^s) = \prod_f F(Y^f),$$

where

$$F(Y^f) = \prod_{\gamma} N(y'_{1\gamma}, \dots, y'_{p_f\gamma}; (\sigma_f)).$$

The proof is almost identical with the proof of Corollary III of which this corollary is an extension.

It is remarked that if the statistics, the limiting distributions of which are desired, are associated with the normal distribution, as are most statistics studied, then Corollary IV may not be the best tool to use. This is a consequence of the fact that such statistics are generally expressible as functions of uncorrelated random variables and hence are more simply discussed, using Corollary I.

**3. Limiting distributions of quadratic and bilinear forms.** We first assume the coefficients of the forms to be constants. For each set of values of  $i, j$ , and  $n$ , the matrix of the bilinear form with coefficients which are real numbers,

$$(3.1) \quad b_{ij}^n = \sum_{\mu, \nu} a_{\mu\nu n} x_{i\mu} x_{j\nu},$$

will be denoted by  $A_n$ , and the rank of  $A_n$  will be denoted by  $m$ . The maximum of the absolute values of the elements of  $A_n$  will be denoted by  $b_n$ . We shall assume that there exists an orthogonal transformation,

$$(3.2) \quad y_{i\mu n} = \sum_{\nu} c_{\mu\nu n} x_{i\nu},$$

of  $x_1, \dots, x_{in}$  such that

$$(3.3) \quad b_{ij}^n = \sum_{\mu} \lambda_{\mu} y_{i\mu n} y_{j\mu n},$$

where the coefficients  $\lambda_{\mu}$  are non-negative.<sup>15</sup>

**LEMMA I.** *If  $d_n$  is the maximum of the absolute values of the elements  $c_{\mu\nu n}$  then a necessary and sufficient condition that  $\lim b_n = 0$  is  $\lim d_n = 0$ .*

<sup>15</sup> Our theorems will not be applicable if some of the  $\lambda_{\mu}$  are negative and some are positive. However if all the  $\lambda_{\mu}$  are non-positive then the theorems will remain true.

PROOF: From (3.1) it follows that

$$a_{\mu\nu n} = \sum_i \lambda_i c_{i\mu n} c_{i\nu n}.$$

Hence,  $b_n \geq a_{\mu\mu n} \geq \lambda_\mu c_{\mu\mu n}^2$  and  $|a_{\mu\nu n}| \leq d_n^2 (\sum_i \lambda_i)$ . The remainder of the proof is obvious.

The following theorem will be the basis for a large sample analogue of Wishart's distribution.

THEOREM III. Let  $\mathcal{H}_p$  be true. Then, a sufficient condition that

$$\lim_{n \rightarrow \infty} F(Y_n) = \prod_{\gamma} N(y_{1\gamma}, \dots, y_{p\gamma}; (\sigma)),$$

where  $b_i^n = \sum_j \lambda_j y_{ij n} y_{jn}$  is  $\lim_{n \rightarrow \infty} b_n = 0$ .

PROOF. According to Lemma I, the fact that  $\lim_{n \rightarrow \infty} b_n = 0$ , implies that  $\lim_{n \rightarrow \infty} d_n = 0$ . The  $y_{i\gamma n}$  are such that  $\mathcal{C}$  is true. Hence the hypotheses of Corollary I are satisfied and the theorem is proved.

Before stating the corollary to Theorem III, we shall prove an obvious lemma which is of constant service.

LEMMA II. Let  $\lim_{n \rightarrow \infty} F(X_n) = F(X)$  at all points of continuity of  $F(X)$ , and let

$$g_{1n} = g_1(x_{1n}, \dots, x_{pn}), \dots, g_{kn} = g_k(x_{1n}, \dots, x_{pn})$$

be Borel measurable functions of their indicated variables for each value of  $n$ , ( $p \geq k$ ), defined on  $R^p$ .

Then

$$\lim F(g_{1n}, \dots, g_{kn}) = F(g_1, \dots, g_k)$$

at all points of continuity of  $F(g_1, \dots, g_k)$ , where  $g_\alpha = g_\alpha(x_1, \dots, x_p)$ .

PROOF. By (2.1), we have

$$(3.4) \quad E[e^{i \sum_{\alpha} t_{\alpha} g_{\alpha}(x_{1n}, \dots, x_{pn})}] = E[e^{i \sum_{\alpha} t_{\alpha} g_{\alpha}}],$$

where since  $g_\alpha(x_1, \dots, x_p)$  is a Borel measurable function of  $x_1, \dots, x_p$  we know that  $g_{1n}, \dots, g_{kn}$  have a joint distribution function  $F(g_{1n}, \dots, g_{kn})$ . Then, since  $\lim_{n \rightarrow \infty} F(X_n) = F(X)$  at all points of continuity of  $F(X)$  we have<sup>17</sup>

$$\lim_{n \rightarrow \infty} E[e^{i \sum_{\alpha} t_{\alpha} g_{\alpha}(x_{1n}, \dots, x_{pn})}] = E[e^{i \sum_{\alpha} t_{\alpha} g_{\alpha}(x_1, \dots, x_p)}]$$

uniformly in every  $t_1, \dots, t_p$  interval since

$$\begin{aligned} & |E[e^{i \sum_{\alpha} t_{\alpha} g_{\alpha}(x_{1n}, \dots, x_{pn})}] - E[e^{i \sum_{\alpha} t_{\alpha} g_{\alpha}(x_1, \dots, x_p)}]| \\ & \leq \int |dF_n(X_1, \dots, X_p) - F(X_1, \dots, X_p)|, \end{aligned}$$

<sup>17</sup> See Cramer, [3, p. 30] and "Additional Note" at the end of the book.

where  $F_n(X_1, \dots, X_p)$  stands for  $F(X_{1n}, \dots, X_{pn})$ , when  $X_i$  and  $X_{in}$  have the same numerical values. It follows from (3.4), that

$$\lim_{n \rightarrow \infty} E[e^{i \sum_{k=1}^p t_k g_{kn}}] = E[e^{i \sum_{k=1}^p t_k g_k}]$$

uniformly in every  $t_1, \dots, t_p$  interval, and consequently

$$\lim_{n \rightarrow \infty} F(g_{1n}, \dots, g_{pn}) = F(g_1, \dots, g_p)$$

at all points of continuity of  $F(g_1, \dots, g_p)$ .

The real valued function  $G_d(x; n, c)$  will be defined by the equations

$$G_d(0; 0, c) = 1, \quad (-\infty < c < \infty),$$

$$G_d(x; n, c) = [\Gamma(\frac{1}{2}n)]^{-1} (2c)^{-1n} x^{1n-1} \exp\left[-\frac{x}{2c}\right], \quad (0 < x < \infty; c > 0; n > 0),$$

and  $G_d(x; n, c) = 0$  otherwise. The function  $G(x; n, c)$  will be defined by the equation

$$G(x; n, c) = \int_0^x G_d(t; n, c) dt.$$

The real valued function  $G_d(x_{11}, x_{12}, \dots, x_{pp}; n, (\sigma))$  will be defined by the equations

$$G_d(0; \dots, 0; p-1, (\sigma)) = 1$$

$$G_d(x_{11}, \dots, x_{pp}; n; (\sigma)) = (2\pi)^{-\frac{1}{2}p(p-1)} \sigma^{-1n} \cdot \left[ \prod_i \Gamma(\frac{1}{2}(n-i+1)) \right]^{-1} \cdot |x|^{\frac{1}{2}(n-p+1)-1}$$

$$\cdot \exp\left[-\frac{1}{2} \sum_{i,j} \sigma^{ij} x_{ij}\right], \quad (0 < x_{ii} < \infty; x_{ij}^2 \leq x_{ii} x_{jj}); (\sigma) \text{ is positive definite},$$

where  $|x|$  is the determinant  $|x_{ij}|$  and  $G_d(x_{11}, \dots, x_{pp}; n, (\sigma)) = 0$  otherwise. The function  $G(x_{11}, \dots, x_{pp}; n, (\sigma))$  will be defined by the equation

$$G(x_{11}, \dots, x_{pp}; n, (\sigma)) = \int_{-\infty}^{x_{pp}} \dots \int_{-\infty}^{x_{11}} G_d(t_{11}, \dots, t_{pp}; n, (\sigma)) dt_{11} dt_{12} \dots dt_{pp}.$$

We can now state the limiting distribution analogue of Wishart's distribution.

**COROLLARY V.** *If  $\mathcal{K}_p$  is true, if  $\lambda_i = 1$ , and if  $m \geq p$  then*

$$\lim_{n \rightarrow \infty} F(b_{11}^n, b_{12}^n, \dots, b_{pp}^n) = G(b_{11}, \dots, b_{pp}; m, (\sigma)).$$

**PROOF.** The conditions of Theorem III and Lemma II are satisfied.

Obviously for fixed  $i$ , the limiting distribution of  $b_{ii}^n$  is  $G(b; m, \sigma_{ii})$ , and if  $i \neq j$ , the limiting distribution of  $b_{ij}^n/m$  is the distribution of the covariance of  $x_i$  and  $x_j$  in a sample of  $m$  independent pairs of observations.<sup>18</sup>

<sup>18</sup> See Wishart and Bartlett, [1, p. 266].

We proceed to the analogue for limiting distributions of one of our generalizations of the Fisher-Cochran theorem. It is first desirable to give some additional definitions.

We consider the bilinear forms

$$(3.5) \quad b_{ij\alpha}^n = \sum_{\mu, \nu} a_{\mu\nu\alpha}^n x_{i\mu} x_{j\nu}$$

with real coefficients, and we denote the matrix of  $b_{ij\alpha}^n$  by  $A_\alpha^n$ . The rank of  $A_\alpha^n$  is  $m_\alpha$ , and the rank of  $A_\alpha^n$  is  $m_{k\alpha}$ . If the maximum of the absolute values of the elements of  $A_\alpha^{k-1}$ ,  $\dots$ ,  $A_\alpha^{k-1}$  is  $b_n$ , and if there exists an orthogonal transformation,

$$(3.6) \quad y_{i\alpha n} = \sum_j c_{j\mu\nu n} x_{i\mu},$$

of  $x_{i1}, \dots, x_{in}$  such that

$$b_{ij\alpha}^n = \sum_i \lambda_i y_{i\alpha n} y_{j\alpha n},$$

where  $\delta$  assumes all integral values from  $m_1 + \dots + m_{\alpha-1} + 1$  through  $m_1 + \dots + m_\alpha$  and  $\lambda_i$  is non-negative, then it is easy to prove, as in Lemma I, that a necessary and sufficient condition that  $\lim_{n \rightarrow \infty} b_n = 0$  is  $\lim_{n \rightarrow \infty} d_n = 0$ , where  $d_n$  is the maximum of the absolute values of the elements  $c_{\mu\nu n}$ .

LEMMA III. Let  $m = m_1 + \dots + m_{k-1}$  and let

$$(3.7) \quad \sum_\alpha b_{ij\alpha}^n = \sum_j x_{i\mu} x_{j\nu}.$$

Then, a necessary and sufficient condition that

$$b_{ij\alpha}^n = \sum_i y_{i\alpha n} y_{j\alpha n},$$

where the real linear functions,  $y_{i\alpha n}$ , of  $x_{i1}, \dots, x_{in}$  are given by (3.6), the linear functions (3.6) not now being assumed to be orthogonal, is

$$m_{k\alpha} = n - m.$$

Furthermore, the functions (3.6) are orthogonal.

The proof of this lemma for the case  $p = 1$  is given in [16]. The procedure to follow in extending the lemma to the cases where  $p > 1$ , is given in [15, p. 473]. It is noted that this lemma is more general than the lemma in [15] inasmuch as we show that the orthogonality of the transformation is a consequence of our hypotheses and not one of the hypotheses.<sup>19</sup>

<sup>19</sup> It is noted, however, that the increase in generality affects only the necessity not the sufficiency of the theorem.



THEOREM IV. Let  $\mathcal{K}_p$ , (3.7) and (3.8) be true for all values of  $n$ , and suppose that  $\lim_{n \rightarrow \infty} b_n = 0$ . Then

$$\lim_{n \rightarrow \infty} F(y_n) = \prod_{\gamma} N(y_{1\gamma}, \dots, y_{p\gamma}; (\sigma)),$$

where  $b_{i\alpha}^n = \sum_j y_{jn} y_{i\alpha n}$ .

The proof is omitted.

COROLLARY VI. If the hypotheses of Theorem IV are assumed, and if  $m_p \geq p$ ; ( $\beta = 1, \dots, h$ ;  $h < k$ ), then

$$\begin{aligned} \lim_{n \rightarrow \infty} F(b_{111}^n, \dots, b_{pph}^n, y_{1h+1n}, \dots, y_{pmn}) \\ = \prod_{\gamma=1}^h G(b_{11\gamma}, \dots, b_{pp\gamma}; m_{\gamma}, (\sigma)) \cdot \prod_{\gamma=h+1}^m N(y_{1\gamma}, \dots, y_{p\gamma}; (\sigma)). \end{aligned}$$

If  $p = 1$  in Theorem IV and Corollary VI, we have the large sample analogue of the Fisher-Cochran theorem.

We now discuss limiting distributions of random variables which are bilinear and quadratic forms in one set of chance variables for fixed values of other random variables. We consider the coefficients  $a_{\mu\nu n}$  and  $a_{\mu\nu n}^{\alpha}$  of  $b_{i\alpha}^n$  and  $b_{i\alpha}^n$  to be random variables. Hence the matrices  $A_n$  and  $A_n^{\alpha}$  are random matrices.

To be more explicit, let  $X_1^f, X_2^f, \dots$  be a sequence of random vectors, the random vector  $X_n^f$  having  $p_f$  components  $x_{1n}^f, \dots, x_{p_f n}^f$ , and being defined on  $R^{p_f}$ . The set of random vectors  $X_n^f$  and  $Z_1, \dots, Z_{i_n}$  will be assumed to be independent.

For each value of  $f$  the coefficients of the bilinear forms

$$(3.9) \quad b_{i\alpha f}^n = \sum_{\mu, \nu=1}^{n_f} a_{\mu\nu\alpha f}^n x_{i\mu}^f x_{i\nu}^f, \quad (i, j = 1, \dots, p_f; \alpha = 1, \dots, k_f)$$

will be assumed to be Borel measurable functions of the random vectors  $X_{\mu}^1, \dots, X_{\mu}^{f-1}$  and  $Z_1, \dots, Z_{i_n}$ .

The matrix of  $b_{i\alpha f}^n$  is denoted by  $A_{n_f}^{\alpha f}$ . The rank of  $A_{n_f}^{\beta f}$  is  $m_{\beta f}$ , and the rank of  $A_{n_f}^{k_f f}$  is  $m_{k_f n_f}$  for all sets of values of the  $a_{\mu\nu\alpha f}^n$  except, perhaps, on a set  $E_{n_f}$  which is such that  $\lim_{n_f \rightarrow \infty} P(E_{n_f}) = 0$ .

Let the function  $b(A_{n_f}^{\beta f})$  be defined as follows:

For each set of values of the  $X_n^f$  and  $Z$  let  $b(A_{n_f}^{\beta f})$  be the maximum of the absolute values of the elements of  $A_{n_f}^{\beta f}$ . We shall denote  $b(A_{n_f}^{\beta f})$  by  $b_{n_f}^{\beta f}$ . Obviously,  $b_{n_f}^{\beta f}$  is a Borel measurable function of  $X_n^f$  and  $Z$ . Hence

$$b_{n_f}^{\beta f} = b(A_{n_f}^{\beta f})$$

is a random variable defined on  $W \times R^{n_1 p_1 + \dots + n_p p_p}$ .

For each value of  $f$ , and for almost all sets of fixed values of the  $X_\mu^h$ , ( $h = 1, \dots, f-1$ ), we shall assume that there exists an orthogonal transformation,

$$(3.10) \quad y'_{i\mu n_f} = \sum_\nu c'_{\mu\nu n_f} x'_{i\nu},$$

of  $x'_{i1}, \dots, x'_{in_f}$ , such that<sup>20</sup>

$$(3.11) \quad b_{i f a f}^{n_f} = \sum_\lambda y'_{i\lambda n_f} y'_{f\lambda n_f},$$

where  $\lambda$  assumes all integral values from  $m_{i f} + \dots + m_{a-1 f} + 1$  through  $m_{i f} + \dots + m_{a f}$ . The coefficients  $c'_{\mu\nu n_f}$  of the linear forms (3.10) are real single valued Borel measurable functions of the coefficients  $a'_{\mu\nu a f}$  of the bilinear forms (3.9) for fixed values of the  $X_\mu^h$  and  $Z^n$ . Let  $c'_{\mu\nu n_f}$  be the same function of the functions  $a'_{\mu\nu a f}$  that  $c'_{\mu\nu n_f}$  is of the coefficients of the bilinear forms having constant coefficients. Furthermore, let  $d'_{n_f}$  be the same function of the matrix  $C'_{n_f} = \|c'_{\mu\nu n_f}\|$  where  $m = m_{i f} + \dots + m_{k f-1 f}$ , that  $b_{n_f}^{a f}$  is of  $A_{m f}^{a f}$ .

LEMMA IV. A necessary and sufficient condition that  $b_{n_f}'$  converge in probability to zero as  $n$  increases is that  $d_{n_f}'$  converge in probability to zero as  $n$  increases.

PROOF. Since

$$\sum_{\beta=1}^{k_f-1} a_{\mu\beta f}^{n_f} = \sum_\lambda c'_{\lambda\mu n_f} c'_{\lambda\beta n_f},$$

we have

$$(k_f - 1)b_{n_f}' \geq \sum_{\beta=1}^{k_f-1} a_{\mu\beta f}^{n_f} \geq [c'_{\lambda\mu n_f}]^2$$

and

$$|a_{\mu\beta f}^{n_f}| \leq \left\{ \sum_\lambda [c'_{\lambda\mu n_f}]^2 \cdot \sum_\lambda [c'_{\lambda\beta n_f}]^2 \right\}^{\frac{1}{2}} \leq m_{a f} [d_{n_f}']^2,$$

where  $\lambda$  assumes all integral values from  $m_{i f} + \dots + m_{a-1 f} + 1$  through  $m_{i f} + \dots + m_{a f}$ . The remainder of the proof is obvious.

In proving Theorem V we shall use a generalization of Lemma III which is proved in [15, p. 473].

THEOREM V. Let  $\mathcal{K}_{p_1}^1 \dots \mathcal{K}_{p_r}^r$  be true, and suppose that

$$\sum_a b_{i f a f}^{n_f} = \sum_{i=1}^{n_f} x'_i x'_{i f}.$$

Then, if  $b_{n_f}'$  converges in probability to zero as  $n$  increases and if  $m_f = n_f - m_{k f n_f}$  for all values of  $n_f$ , it follows that

$$\lim_{n_1, \dots, n_r \rightarrow \infty} F(y_{1n_1}^1, \dots, y_{p_r m_r n_r}^r) = \prod_f \prod_{\gamma=1}^{m_f} N(y'_{1\gamma}, \dots, y'_{p_f \gamma}; (\sigma')).$$

The proof is omitted.

<sup>20</sup> It is not necessary that the  $\lambda$ 's be set equal to one as in (3.11). It is only somewhat easier to state the results.

COROLLARY VII. If  $m_{\alpha f} \geq p_f$ , then

$$\lim_{n_1, \dots, n_s \rightarrow \infty} F(b_{1111}^{n_1}, \dots, b_{p_s p_s k_s - 1}^{n_s}) = \prod_f \prod_{j=1}^{k_f-1} G(b_{11j f}, \dots, b_{p_f p_f j f}; m_{\alpha f}, (\sigma^f)).$$

The proof is omitted.

Finally, let us assume that the vectors  $X'_i$ , for fixed  $i$  are uncorrelated and for fixed  $f$  are independent. By that, we shall mean that  $E(x'_i x'_j) = \sigma'_{ij} \delta_{if}$  and that for all  $n$  the set of random vectors  $X'_i$  are independent for the same or different superscripts providing the subscripts are all different. Let us also assume that the coefficients of the forms (3.9) are real numbers. Thus we have weakened the hypotheses of Theorem V concerning the random vectors, and we have strengthened the hypotheses of Theorem V concerning the forms (3.9). Inasmuch as we are generally concerned with the limiting distributions of statistics which occur in the analysis of the normal distribution, and many such statistics have been shown to be invariant under transformations into uncorrelated random variables,<sup>21</sup> Theorem VI and Corollary VIII will often be applicable.

THEOREM VI. The statement of Theorem V is repeated.

COROLLARY VIII. The statement of Corollary VII is repeated.

Another extension of these theorems may be obtained by allowing all the  $n_i$  to be equal, i.e.  $n_1 = \dots = n_s = n$ , and by putting conditions on the forms (3.9) which enable us to say that for fixed  $i, f, \mu$  and  $n$ , the set of random variables  $c_{\mu n} x'_i$  are independently distributed. Theorem I could then be used to obtain a very general result. However, except for the case dealt with above, the condition of independence appears to be rather restrictive, and the theorem is omitted.

**4. Applications.** We first state the strong law of large numbers and a lemma which is very useful in the discussion of limiting distributions.

A sequence of random variables  $X_1, \dots$  will be said to converge with probability one<sup>22</sup> to a random variable  $X$  if

$$\lim_{n \rightarrow \infty} P\{|X_n - X| < \epsilon, |X_{n+1} - X| < \epsilon, \dots, |X_{n+p} - X| < \epsilon\} = 1$$

for every value of  $p \geq 0$ , uniformly in  $p$  for every positive number  $\epsilon$ . Upon setting  $p = 1$ , it is seen that convergence with probability one implies convergence in probability.

The strong law of large numbers<sup>23</sup> asserts that if the independent random variables  $X, X_1, \dots$  all have the same distribution function, and if  $E(X)$  is finite, then the sequence of arithmetic means  $\frac{1}{n} \sum_{i=1}^n X_i$  converges with probability one to  $E(X)$ .

<sup>21</sup> The regression transformation which yields the uncorrelated variables will be found in [15, p. 476, (3.2)].

<sup>22</sup> See Doob [4, p. 163], and Frechet, [9, p. 228].

<sup>23</sup> See Doob [4, p. 163], and Frechet, [9, p. 250]. A complete proof is given by Frechet.

Hence, if  $E(x_{i\nu}) = 0$  and if  $\sigma_{ij}$  is finite, then  $\frac{1}{n} \sum x_{i\nu} x_{j\nu} = s'_{ijn}$  converges with probability one to  $\sigma_{ij}$ . Since  $\sum (x_{i\nu} - \bar{x}_{in})(x_{j\nu} - \bar{x}_{jn}) = \sum x_{i\nu} x_{j\nu} - n\bar{x}_{in}\bar{x}_{jn}$  where  $\bar{x}_{in}$  is the arithmetic mean of  $x_{i1}, \dots, x_{in}$ , and since  $\bar{x}_{in}$  converges with probability one to zero, it follows that  $s_{ijn} = s'_{ijn} - \bar{x}_{in}\bar{x}_{jn}$  converges with probability one to  $\sigma_{ij}$ . It is, of course, assumed that the random variables  $x_{i\nu}, x_{j\nu}$  have the same joint distribution function for all values of  $\nu$ , and that the random vectors  $X_1, \dots$  are independently distributed. The process of the reduction of  $s_{ijn}$  to  $s'_{ijn}$  in the limit, is an example of the possible uses of:

**LEMMA V.** *If  $\varphi(t_1, \dots, t_p)$  is a continuous function of  $t_1, \dots, t_p$ , and if the sequence of random variables  $x_{in}$  converges in probability, (with probability one) to  $x_i$  which may be a random variable or a constant, then the sequence of random variables  $\varphi(x_{1n}, \dots, x_{pn})$  converges in probability (with probability one) to  $\varphi(x_1, \dots, x_p)$ , where some or all of the  $x$ 's may be constants. If  $x_1, \dots, x_p$  are constants then  $\varphi(t_1, \dots, t_p)$  need only be continuous in the neighborhood of  $x_1, \dots, x_p$  and Borel measurable.*

For a proof of part of this lemma which may be extended to yield the entire proof, see, Frechet, [9, p. 178].

Using Lemma V it is easy to see that the coefficients  $r_n$  of least squares equations converge with probability one to their  $\beta$  values, where the  $\beta$  value is obtained by substituting  $\sigma_{ij}$  for  $s_{ijn}$  in the expression for  $r_n$  assuming, of course, independent random vectors which have the same distribution functions.

Since problems in the analysis of variance may be interpreted as problems in least squares the above comments and Lemma V will generally make it possible, when determining limiting distributions, to consider the statistics to be functions of deviations from "true" mean functions rather than "sample" mean functions.

We shall discuss, briefly, four applications of these results.

(a). *The limiting distribution of the regression coefficient.* Let  $r_n$ , the "sample" regression coefficient, be defined by the equation

$$r_n = \frac{\sum x_{i\nu} x_{j\nu}}{\sum x_{i\nu}^2}$$

where  $x_{i\nu}$  and  $x_{j\nu}$  are deviations from arithmetic means. If the random vectors  $(x_{i\nu}, x_{j\nu})$  are independently distributed for fixed  $i, j$ , with the same distribution functions, and if  $E(x_{i\nu}) = E(x_{j\nu}) = 0$ ,  $E(x_{i\nu} x_{j\nu}) = \sigma_{ij}$ , then it follows from the strong law of large numbers that  $\sum x_{i\nu} x_{j\nu} / n$  converges to  $\sigma_{ij}$  with probability one, and from the Laplace-Liapounoff theorem that  $\sum x_{i\nu} x_{j\nu} / \sqrt{n}$  has a normal limiting distribution with mean  $\sigma_{ij}$  and variance  $E\{x_{i\nu} x_{j\nu} - \sigma_{ij}\}^2$ . Hence, by Lemma V,  $\sqrt{n} \left( r_n - \frac{\sigma_{ij}}{\sigma_{ii}} \right)$  has a normal limiting distribution with mean zero and variance  $\lim_{n \rightarrow \infty} E \left\{ n \left( r_n - \frac{\sigma_{ij}}{\sigma_{ii}} \right)^2 \right\}$  unless that limit does not exist.

If the  $x_{i\nu}$  are not random variables then, in order to apply Corollary I with  $p = 1$ , it is necessary that

$$(4.1) \quad \lim_{n \rightarrow \infty} \frac{x_{in}}{(\sum_{\nu} x_{i\nu}^2)^{1/2}} = 0.$$

In that case, the limiting distribution of  $(\sum_{\nu} x_{i\nu}^2)^{1/2} \cdot r_n$  is normal with zero mean and variance  $\sigma_{ij}$ . If (4.1) is not satisfied then there is no assurance, unless the  $x_{i\nu}$  are normally distributed, that the limiting distribution of  $(\sum_{\nu} x_{i\nu}^2)^{1/2} r_n$  is normal.

(b). *The limiting distribution of the analysis of variance ratio.* The tests of significance which occur in the analysis of variance depend on the ratio of two quadratic forms,  $q_{1n}$  and  $q_{2n}$ , the denominator  $q_{2n}$  having rank (or degrees of freedom)  $m_{2n}$  increasing with  $n$ , and the numerator  $q_{1n}$  having rank  $m_1$  not changing with  $n$ , i.e.,

$$v_n = \frac{q_{1n} m_{2n}}{q_{2n} m_1},$$

where  $q_{1n} + q_{2n} + q_{3n} = \sum_{\nu} x_{i\nu}^2$  and  $q_{3n}$  is a quadratic form of rank  $m_{3n}$  which will be identically zero if  $n = m_1 + m_{2n}$ . Since<sup>24</sup>  $q_{2n}$  is expressible as the variance of  $x$  about a least squares equation it follows from the previous discussion and Lemma IV that  $\frac{q_{2n}}{m_{2n}}$  converges with probability one to  $\sigma^2$  under the assumptions that the  $x_{i\nu}$  are independently distributed with zero means and variances  $\sigma^2$ . Hence the limiting distribution of  $v_n$  will depend only on the limiting distribution of  $q_{1n}$  and it will consequently be necessary to consider only the matrix of  $q_{1n}$ , in order to apply Corollary VI with  $p = 1$ . For example,<sup>25</sup> if there are  $pn$  independently distributed random variables  $x_{i\nu}$  with zero means and variances  $\sigma^2$  arranged in  $p$  blocks of  $n$  random variables each, then

$$\sum_{i,\nu} (x_{i\nu} - \bar{x})^2 = n \sum_i (\bar{x}_{in} - \bar{x}_n)^2 + \sum_{i,\nu} (x_{i\nu} - \bar{x}_{in})^2,$$

where  $\bar{x}_{in}$  is the arithmetic mean of  $x_{i1}, \dots, x_{in}$  and  $\bar{x}_n$  is the arithmetic mean of all the  $x_{i\nu}$ . Then

$$q_{1n} = n \sum_i (\bar{x}_{in} - \bar{x}_n)^2,$$

$$q_{2n} = \sum_{i,\nu} (x_{i\nu} - \bar{x}_{in})^2,$$

$$m_1 = p - 1,$$

$$m_{2n} = p(n - 1)$$

<sup>24</sup> This has been proved by Kolodziejczyk, [12, p. 161].

<sup>25</sup> Other schemes are given in Fisher, [8].

and the matrix of  $q_{1n}$  may be obtained by substituting for the  $x_{in}$  and  $\bar{x}_n$ . In this case it is sufficient to express  $q_{1n}$  as  $\sum_{i,j} a_{ij} S_i S_j$  where  $S_i = \sum_j x_{ij}$ ,  $a_{ii} = (p-1)/pn$ , and,  $f, i \neq j$ ,  $a_{ij} = -1/pn$ , to see that the condition that the maximum of the absolute values of the elements of the matrix of  $q_{1n}$  approaches zero as  $n$  increases. Hence, if the  $x_{in}$  satisfy the condition  $\mathfrak{L}$ , the limiting distribution of  $m_1 v_n$  is  $G(v; p-1, 1)$ .

Clearly, if only the rank of  $q_{2n}$  increases as  $n$  increases, the rank  $m_{2n}$  of  $q_{2n}$  being constant and if the maximum of the absolute values of the elements of the matrix of  $q_{2n}$  also approaches zero as  $n$  increases, then  $v_n$  will have a limiting distribution which is the analysis of variance distribution, and the limiting distribution of  $\frac{q_{1n}}{q_{1n} + q_{2n}}$  will be the correlation ratio distribution.

(c). *Periodogram analysis.* We need only remark that the linear functions which are used in the analysis of the Schuster periodogram<sup>26</sup> meet all the requirements of Corollary I if the  $x_i$  are independently distributed with zero means and constant variances and satisfy the condition  $\mathfrak{L}$ . Consequently the large sample theory of the Schuster periodogram is the same for non-normal as it is for normal distributions.

(d). *Multivariate analysis.* We shall assume that the random vectors  $X_1, \dots, (X_p$  has components  $x_{1p}, \dots, x_{pn})$ , are independently distributed, that (2.3) and (2.4) are satisfied, and that the condition  $\mathfrak{L}_p$  is satisfied. For any fixed  $n$  and  $\alpha$  we shall call the determinant  $D_\alpha^n$  of the forms (3.5) a generalized sum of squares, and the determinant  $V_\alpha^n$  of the elements  $b_{ij\alpha}^n/m_\alpha$  a generalized variance. We shall say that  $D_\beta^n$  and  $V_\beta^n$  have rank  $m_\beta$  and that  $D_k^n$  and  $V_k^n$  have rank  $n_{kn}$ . If  $m_\beta$  is constant, and if (3.7) and (3.8) are true then clearly the limiting distribution of  $D_\beta^n$  is the distribution of the generalized variance of  $m_\beta$  vector observations<sup>27</sup> from a normal distribution, with zero means and covariance parameters  $\sigma_{ij}$ . Under the same conditions, the limiting distribution of  $D_\beta^n/V_k^n$  is the distribution of the generalized variance of  $m_\beta$  vector observations from a normal distribution with zero means and covariance parameters  $\delta_{ij}$ . Many other similar limiting distributions are immediately derivable.

Before completing our discussion of the limiting distributions of statistics occurring in multivariate analysis, we shall state a theorem on limiting distributions which is an obvious generalization of a theorem of Doob, [4, p. 166].

Suppose that the random variables  $g(n)X_{1n}, \dots, g(n)X_{pn}$  have a distribution function  $F(g(n)X_{1n}, \dots, g(n)X_{pn})$  which is such that

$$\lim_{n \rightarrow \infty} F(g(n)X_{1n}, \dots, g(n)X_{pn}) = F(X_1, \dots, X_p),$$

where  $F(X_1, \dots, X_p)$  is a continuous distribution function, and suppose that  $X_{in}$  converges in probability to the real number  $\xi_i$ . For example, if  $\bar{x}_n =$

<sup>26</sup> The theory of the Schuster periodogram is given by Fisher [7].

<sup>27</sup> See Wilks, [18, p. 476] or Madow, [15, pp. 481, 484].

$\sum_j x_j/n$  where  $E(x_r) = 0$ ,  $E(x_r^2) = 1$ , and  $\mathfrak{L}$  is satisfied, then  $\bar{x}_n$  converges to zero with probability one, and  $\sqrt{n} \bar{x}_n$  has a limiting distribution which is normal with zero mean and unit variance, i.e.

$$\lim_{n \rightarrow \infty} |P\{\sqrt{n}\bar{x}_n < x\} - N(X; 1)| = 0.$$

**THEOREM VII.** Let  $\varphi_f(t_1, \dots, t_p)$  be a function of  $t_1, \dots, t_p$  defined in a neighborhood  $N$  of  $\xi_1, \dots, \xi_p$  which, together with its  $(k_f + 1)$ -th partial derivatives is continuous in  $N$ . Suppose that  $k$  is the least value of  $r_f$  such that the random variables<sup>28</sup>

$$[g(n)]^{r_f} \left[ \sum_i (x_{in} - \xi_i) \cdot \frac{\partial \varphi_f(\xi_1, \dots, \xi_p)}{\partial \xi_i} \right]^{(r_f)}$$

have a joint limiting distribution function  $D(x_1, \dots, x_s)$ . Then the random variables  $[g(n)]^{k_f} [\varphi_f(x_{1n}, \dots, x_{pn}) - \varphi_f(\xi_1, \dots, \xi_p)]$  have a joint limiting distribution which is given by  $D(x_1, \dots, x_s)$ . The value  $k_f$  is greater than or equal to the minimum value for which not all the partial derivatives of order  $k_f$  vanish at  $\xi_1, \dots, \xi_p$ .

The proof is almost word for word that of Doob, the only difference being the removal of the specializing words.

We now consider the limiting distribution of the ratio of generalized sums of squares  $L_n$  which is defined by

$$L_n = \frac{D_k^n}{D_{k+1}^n},$$

where  $D_{k+1}^n$  is the determinant of the forms  $b_{ij}^n + b_{i+1,j}^n = b_{ij}^n + b_{i,j+1}^n$ . It has been shown that<sup>29</sup>

$$L_n = \prod_i \frac{Y_{ik}^n}{Y_{ik+1}^n},$$

where  $Y_{ij}^n$ , ( $j = k, k + 1$ ), is a ratio of generalized sums of squares

$$Y_{ij}^n = \frac{|b_{rsj}^n|}{|b_{uvj}^n|}, \quad (r, s = 1, \dots, i; u, v = 1, \dots, i - 1; b_{00j}^n = 1).$$

Since  $Y_{ij}^n/m_{jn}$  converges with the probability one to  $|\sigma_{rs}|/|\sigma_{uv}|$ , and since, by Corollary VIII the joint limiting distribution of the  $m_{k+1,n} \left(1 - \frac{Y_{ik}^n}{Y_{ik+1}^n}\right)$  is

<sup>28</sup> See Goursat-Hedrick, [10, p. 107] for a statement of the Taylor expansion of functions of several variables, which we use here, by  $\frac{\partial \varphi_f(\xi_1, \dots, \xi_p)}{\partial \xi_i}$  is meant the value of  $\frac{\partial \varphi_f(x_1, \dots, x_p)}{\partial x_i}$  at the point  $\xi_1, \dots, \xi_p$ .

<sup>29</sup> See Madow, [15, p.

$\prod_i G(x_i; m, 1)$  it follows, by Theorem VII, that the joint limiting distribution of the ratios of generalized sums of squares

$$\prod_{k=1}^i \frac{Y_{kk}^n}{Y_{kk+1}^n},$$

is

$$\prod_i G(x_i; im_1, 1)$$

and that the limiting distribution of  $m_{k+1}^{-1}(1 - L_n)$  is<sup>30</sup>

$$G(x; pm_1, 1).$$

In a following paper, these results will be extended to quadratic forms in non-central random variables.

**5. Summary.** In Section 2, Theorem I, we stated a very general form of the Laplace-Liapounoff theorem based on the Lindeberg condition. In four corollaries, this theorem was shown to provide joint limiting distributions for systems of linear forms which are such that the maximum of the absolute values of their coefficients converge to zero with an increase in the size of the sample if the coefficients are constants, and converge in probability to zero with an increase in the size of the sample if the coefficients are themselves random variables. It was shown that under certain conditions functions of several random variables, which are such that each function is a linear function of certain random variables for fixed values of random variables of lower index, also have a normal multivariate limiting distribution.

These results were extended to include limiting distributions of quadratic and bilinear forms in Section 3. The method of extension was to show that necessary and sufficient conditions for the existence of systems of linear forms satisfying the conditions of Section 2 are provided by rather simple conditions, the most important of which is that the greatest of the absolute values of the elements of the matrices of the quadratic and bilinear forms approach zero if the size of the sample increases, the ranks of the forms remaining unaltered. This led to the theorem that quadratic and bilinear forms having such matrices have  $\chi^2$ , or covariance, or Wishart's distribution as limiting distributions. It was then shown, in Theorem IV, that if the rank of the sum of the matrices of the quadratic and bilinear forms is equal to the sum of the ranks of the matrices, and if certain of these ranks do not change as the size of the sample increases, then the system of quadratic and bilinear forms have Wishart's distribution in the limit provided the other conditions are met. These results

<sup>30</sup> A generalization of Wilks' result, [19, p. 323] to the case where the variates are not assumed to have a normal multivariate distribution may readily be obtained.



were then extended in Theorem V to one of the cases occurring when the coefficients of the forms are themselves random variables.

Several simple illustrations of the uses of the methods were given in Section 4. It was shown that the analysis of the variance ratios, and statistics occurring in the theory of multivariate statistical analysis have the same limiting distributions which they would have had if their variables had been normally and independently distributed.

#### REFERENCES

- [1] M. S. BARTLETT AND J. WISHART, "The generalized product moment distribution in a normal system," *Proc. Camb. Phil. Soc.*, Vol. 29 (1933), pp. 260-270.
- [2] W. G. COCHRAN, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance," *Proc. Camb. Phil. Soc.*, Vol. 30 (1934), pp. 178-191.
- [3] H. CRAMER, *Random Variables and Probability Distributions*, *Camb. Tracts in Math. and Math. Physics*, No. 36, London, 1937.
- [4] J. L. DOOB, "The limiting distributions of certain statistics," *Annals of Math. Stat.*, Vol. 6 (1935), pp. 160-169.
- [5] J. L. DOOB, "Stochastic processes with an integral-valued parameter," *Trans. Am. Math. Soc.*, Vol. 44 (1938), pp. 87-150.
- [6] R. A. FISHER, "Applications of 'Student's' distribution," *Metron*, Vol. 5 (1926), pp. 90-104.
- [7] R. A. FISHER, "Tests of significance in harmonic analysis," *Proc. Roy. Soc., (A)*, Vol. 125 (1929), pp. 54-60.
- [8] R. A. FISHER, *Statistical Methods for Research Workers*, 7th ed., Oliver and Boyd, London, 1938.
- [9] M. FRECHET, *Recherches Theoriques Modernes sur la Theorie des Probabilites*, Vol. 1, Gauthier-Villars, Paris, 1937.
- [10] E. GOURSAT, *A Course in Mathematical Analysis*, Vol. 1, translated by E. R. Hedrick, Ginn and Co., New York, 1904.
- [11] A. KOLMOGOROFF, "Grundbegriffe der Wahrscheinlichkeitsrechnung," *Ergebnisse der Mathematik*, Vol. 2, no. 3.
- [12] S. KOLODZIEJCZYK, "On an important class of statistical hypotheses," *Biometrika*, Vol. 27 (1935), pp. 161-190.
- [13] P. LEVY, *Calcul des Probabilites*, Gauthier-Villars, Paris, 1925.
- [14] P. LEVY, *Theorie de l'Addition des Variables Aleatoires*, Gauthier-Villars, Paris, 1937.
- [15] W. G. MADOW, "Contributions to the theory of multivariate statistical analysis," *Trans. Am. Math. Soc.*, Vol. 44 (1938), pp. 454-495.
- [16] W. G. MADOW, "The distribution of quadratic forms in non-central normal random variables," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 100-103.
- [17] S. SAKS, *Theory of the Integral*, 2nd ed., G. E. Stechert and Co., New York, 1937.
- [18] S. S. WILKS, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 472-494.
- [19] S. S. WILKS, "On the Independence of  $k$  sets of normally distributed statistical variables," *Econometrika*, Vol. 3 (1935), pp. 309-326.
- [20] J. WISHART AND M. S. BARTLETT, See [1].

WASHINGTON, D. C.

# ON A TEST WHETHER TWO SAMPLES ARE FROM THE SAME POPULATION<sup>1</sup>

BY A. WALD<sup>2</sup> AND J. WOLFOWITZ

**1. The Problem.**<sup>3</sup> Let  $X$  and  $Y$  be two independent stochastic variables about whose cumulative distribution functions nothing is known except that they are continuous. Let  $x_1, x_2, \dots, x_m$  be a set of  $m$  independent observations on  $X$  and let  $y_1, \dots, y_n$  be a set of  $n$  independent observations on  $Y$ . It is desired to test the hypothesis (the null hypothesis) that the distribution functions of  $X$  and  $Y$  are identical.

An important step in statistical theory was made when "Student" proposed his ratio of mean to standard deviation for a similar purpose. In the problem treated by "Student" the distribution functions were assumed to be of known (normal) form and completely specified by two parameters. It is clear that in the problem to be considered here the distributions cannot be specified by any finite number of parameters.

It might nevertheless be argued that by virtue of the limit theorems of probability theory, "Student's" ratio might be used in our problem for large samples. Such a procedure is open to very serious objections. The population distributions may be of such form (e.g., Cauchy distribution) that the limit theorems do not apply. Furthermore, the distributions of  $X$  and  $Y$  may be radically different and yet have the same first two moments; clearly "Student's" ratio will not distinguish between two such distributions.

The Pearson contingency coefficient is a useful test specifically designed for the problem we are discussing here, but one which also possesses some disadvantages. The location of the class intervals is to a considerable extent arbitrary. In order to use the  $\chi^2$  distribution, the numbers in each class interval must not be small; often this can be done only by having large class intervals, thus entailing a loss of information.

**2. Preliminary remarks.** Denote by  $P\{X < x\}$  the probability of the relation in braces. Let  $f(x)$  and  $g(x)$  be the distribution functions of  $X$  and  $Y$  respectively; e.g.,  $P\{X < x\} = f(x)$ . Throughout this paper we shall assume that  $f(x)$  and  $g(x)$  are continuous.

Let the set of  $m + n$  elements  $x_1, \dots, x_m$  and  $y_1, \dots, y_n$  be arranged in

<sup>1</sup> Presented to the Institute of Mathematical Statistics at Philadelphia, December 27, 1939.

<sup>2</sup> Research under a grant-in-aid from the Carnegie Corporation of New York.

<sup>3</sup> The authors are indebted to Prof. S. S. Wilks for proposing this problem to them.

ascending order of magnitude, and let the sequence be designated by  $Z$ , thus:  $Z = z_1, z_2, \dots, z_{m+n}$ , where  $z_1 < z_2 < \dots < z_{m+n}$ . ( $f(x)$  and  $g(x)$  were assumed to be continuous. Hence the probability is 0 that  $z_i = z_{i+1}$  and therefore we may exclude this case.) Let  $V = v_1, v_2, \dots, v_{m+n}$  be a sequence defined as follows:  $v_i = 0$  if  $z_i$  is a member of the set  $x_1, \dots, x_m$  and  $v_i = 1$  if  $z_i$  is a member of the set  $y_1, \dots, y_n$ . It is easy to show that any statistic  $S$  used to test the null hypothesis should be invariant under any continuous, reciprocally one-to-one transformation of the real axis. That is to say, if  $t' = \varphi(t)$  is any such transformation, then

$$(1) \quad S(x_1, \dots, x_m, y_1, \dots, y_n) \equiv S(\varphi(x_1), \dots, \varphi(x_m), \varphi(y_1), \dots, \varphi(y_n)).$$

The reason for this requirement on  $S$  is the fact that the transformed stochastic variables  $X' = \varphi(X)$  and  $Y' = \varphi(Y)$  are continuous and have identical distributions if and only if  $X$  and  $Y$  have identical distributions. Hence  $S$  must be a function of  $V$  only, with the added restriction that  $S(V) = S(V')$ , where  $V' = v_{m+n}, v_{m+n-1}, \dots, v_1$ . For if  $S$  were a function of  $x_1, \dots, x_m, y_1, \dots, y_n$  which cannot be expressed as a function of  $V$  alone, then there exists a continuous reciprocally one-to-one transformation  $t' = \varphi(t)$  such that (1) is not true. On the other hand, any continuous reciprocally one-to-one transformation of the entire line into itself is monotonic and hence either leaves  $V$  invariant or else transforms it into  $V'$ .

**3. Previous results.** In an interesting paper on this problem W. R. Thompson [1] proceeds as follows: Let the sets  $x_1, \dots, x_m$  and  $y_1, \dots, y_n$  be ordered in ascending order of magnitude, thus:  $x_{p_1}, x_{p_2}, \dots, x_{p_m}$  and  $y_{p'_1}, y_{p'_2}, \dots, y_{p'_n}$  where  $x_{p_1} < x_{p_2} < \dots < x_{p_m}$  and  $y_{p'_1} < y_{p'_2} < \dots < y_{p'_n}$ . Let  $P\{x_{p_k} < y_{p'_k}\}$  denote the probability of the relation in braces under the null hypothesis ( $f(x) \equiv g(x)$ ). This probability is shown to be independent of  $f(x)$  and the relation

$$(2) \quad P\{x_{p_k} < y_{p'_k}\} = \psi(m, n, k, k')$$

holds, where the right member, which is given explicitly by Thompson, is a function only of the arguments exhibited. To make a test of the null hypothesis with, say, a 5% level of significance, this writer proposes to choose  $k$  and  $k'$  so that  $\psi(m, n, k, k') = .05$ . The test would then consist of noticing whether  $x_{p_k} < y_{p'_k}$  or not. In the former case the null hypothesis is to be considered as disproved.

It is clear that this test cannot be very efficient, ignoring as it does so many of the relations among the observations. Except under certain rather narrow restrictions on the admissible alternatives, for example, that  $g(x) \equiv f(x + c)$ , where  $c$  is an arbitrary constant, the test suffers the further defect of not being "consistent" in a way which will be discussed below. Hence the test suggested by Thompson can scarcely be regarded as a satisfactory solution of the problem. This criticism, of course, does not apply to those sections of Thompson's paper which deal with the question of estimating the so-called normal range.

**4. The statistic  $U$ .** A subsequence  $v_{s+1}, v_{s+2}, \dots, v_{s+r}$  of  $V$  (where  $r$  may also be 1) will be called a "run" if  $v_{s+1} = v_{s+2} = \dots = v_{s+r}$  and if  $v_s \neq v_{s+1}$  when  $s > 0$  and if  $v_{s+r} \neq v_{s+r+1}$  when  $s+r < m+n$ . For example,  $V = 1, 0, 0, 1, 1, 0$  contains the following runs: 1; 0, 0; 1, 1; 0. The statistic<sup>4</sup>  $U$  defined as the number of runs in  $V$  seems a suitable statistic for testing the hypothesis that  $f(x) \equiv g(x)$ . In the event that the latter identity holds, the distribution of  $U$  is independent of  $f(x)$ . A difference between  $f(x)$  and  $g(x)$  tends to decrease  $U$ .  $U$  is consistent in a sense which will be discussed below.

In order to derive the distribution of  $U$  under the null hypothesis, we first note that all the  $\frac{(m+n)!}{m!n!}$  ( $= {}^{m+n}C_m$ ) possible sequences  $V$  have the same probability ( $= \frac{1}{(m+n)!}$ ). To see this, consider the sequence  $V$  where  $v_i = 0$  ( $i = 1, 2, \dots, m$ ) and  $v_i = 1$  ( $i = m+1, m+2, \dots, m+n$ ). Clearly the probability of the sequence is

$$\frac{m(m-1) \dots 1 \cdot n(n-1) \dots 1}{(m+n)(m+n-1) \dots (n+1)n(n-1) \dots 1}.$$

Furthermore, the probability of any other sequence is equal to the product of the factors in the numerator of  $q$  taken in a different order, divided by the product of the factors in the denominator taken in the same order. The quotient is, of course,  $= q$ .

Let  $e_0$  be the number of runs in  $V$  whose elements are 0 and let  $e_1$  be the number of runs whose elements are 1. Obviously  $U = e_0 + e_1$ . Let the runs of each kind be arranged in the ascending order of the indices of the  $v_i$ . Let  $r_{0j}$  be the number of elements 0 in the  $j^{\text{th}}$  run of that kind ( $j = 1, 2, \dots, e_0$ ) and let  $r_{1j'}$  be the number of elements 1 in the  $j'^{\text{th}}$  run of that kind ( $j' = 1, 2, \dots, e_1$ ). The following relations obviously hold:

$$(3) \quad \sum_{j=1}^{e_0} r_{0j} = m,$$

$$(4) \quad \sum_{j'=1}^{e_1} r_{1j'} = n,$$

$$(5) \quad 1 \leq e_0 \leq m, \quad 1 \leq e_1 \leq n,$$

$$(6) \quad |e_0 - e_1| \leq 1.$$

<sup>4</sup>When this paper was already in proof, our attention was called to a paper by W. L. Stevens, entitled "Distribution of groups in a sequence of alternatives," *Annals of Eugenics*, Vol. 9 (1939). There a statistic, which is essentially the  $U$  statistic, is proposed for a problem different from that considered by us and the distribution of  $U$  is obtained in a different manner. However, the application of the  $U$  statistic for the purpose herein described, the proof of consistency and the other results of our paper are not contained in it.

Hence if  $U = 2k$ , then  $e_0 = e_1 = k$ , and if  $U = 2k - 1$ , then either  $e_0 = k$ ,  $e_1 = k - 1$  or  $e_0 = k - 1$ ,  $e_1 = k$ . The element  $v_1$  of  $V$  together with the numbers  $r_{01}, r_{02}, \dots, r_{0e_0}, r_{11}, r_{12}, \dots, r_{1e_1}$ , completely determines the sequence  $V$  whose probability is  $q$ .

Without loss of generality we may assume that  $m \leq n$ . If  $U = 2k$ ,  $1 \leq k \leq m$ ,  $v_1 = 0$ , any two sequences of  $k$  positive numbers each may constitute a sequence of  $r_{01}, \dots, r_{0e_0}, r_{11}, \dots, r_{1e_1}$  provided only that (3) and (4) are satisfied. The number of sequences  $r_{01}, r_{02}, \dots, r_{0k}$  which satisfy (3) is the coefficient of  $a^m$  in the purely formal expansion of

$$(a + a^2 + a^3 + \dots)^k = \left( \frac{a}{1-a} \right)^k$$

and hence is  ${}^{m-1}C_{k-1}$ . Similarly the number of sequences  $r_{11}, r_{12}, \dots, r_{1k}$  which satisfy (4) is found to be  ${}^{n-1}C_{k-1}$ . Bearing in mind the case  $U = 2k$ ,  $v_1 = 1$ , we obtain

$$(7) \quad P\{U = 2k\} = \frac{2({}^{m-1}C_{k-1} \cdot {}^{n-1}C_{k-1})}{m+nC_m}, \quad (k = 1, 2, \dots, m),$$

where the left member denotes the probability of the relation in braces under the null hypothesis. In a similar manner we obtain

$$(8) \quad P = \{U = 2k - 1\} = \frac{({}^{m-1}C_{k-1} \cdot {}^{n-1}C_{k-2} + {}^{m-1}C_{k-2} \cdot {}^{n-1}C_{k-1})}{m+nC_m},$$

$$(k = 2, \dots, m + 1),$$

with the proviso that  ${}^aC_b = 0$  if  $a < b$ .

We shall now briefly indicate a method of obtaining the mean  $E(U)$  and variance  $\sigma^2(U)$  of  $U$ . For example,  $E(U)$  may be obtained by performing several summations of the type

$$(9) \quad \sum_{i=0}^{m-1} i \cdot {}^{m-1}C_i \cdot {}^{n-1}C_i.$$

It is easy to verify that the expression (9) is the term free of  $a$  in the purely formal expansion in  $a$  of:

$$(10) \quad (m-1) \cdot (1+a)^{m-2} \cdot a \cdot \left(1 + \frac{1}{a}\right)^{n-1},$$

and hence is

$$(11) \quad (m-1) \cdot {}^{m+n-3}C_{n-2}.$$

The other summations required for the mean and variance can be carried out in a similar manner. We shall omit these tedious calculations. The results are:

$$(12) \quad E(U) = \frac{2mn}{m+n} + 1,$$

$$(13) \quad \sigma^2(U) = \frac{2mn(2mn - m - n)}{(m+n)^2(m+n-1)}.$$

The critical region for testing the null hypothesis on a level of significance  $\beta$  is given by the inequality  $U < u_0$ , where  $u_0$  is a function of  $m$  and  $n$  such that  $P\{U < u_0\} = \beta$ .

**5. The asymptotic distribution of  $U$ .** Let  $m/n = \alpha$ , a positive constant. Then, as  $m \rightarrow \infty$ ,

$$E(U) \rightarrow \frac{2m}{1+\alpha},$$

$$\sigma^2(U) \rightarrow \frac{4\alpha m}{(1+\alpha)^3}.$$

**THEOREM I.** *If  $t$  is any real number, the probability of the relation  $U < \frac{2m}{1+\alpha} + 2\left[\frac{\alpha m}{(1+\alpha)^3}\right]^{\frac{1}{2}}t$  converges uniformly in  $t$  to*

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}w^2} dw$$

as  $m \rightarrow \infty$ .

The proof of this theorem is essentially the same as the classical proof that the binomial law converges to the normal distribution (see, for example, Fréchet [2], p. 89) and it will be unnecessary to give the details. Since the asymptotic distribution of the subpopulation of even  $U$  is the same as that of odd  $U$ , it will be sufficient to consider only the right member of (7). Let  $m' = m - 1$ ,  $n' = n - 1$ , and  $k' = k - 1$ . We make the substitution

$$(14) \quad w = \frac{k' - \frac{m'}{1+\alpha'}}{\sqrt{\frac{m'}{1+\alpha'}}}, \quad \text{where} \quad \alpha' = \frac{m'}{n'},$$

$$(15) \quad dw = \frac{1}{\sqrt{m'}},$$

and evaluate the factorials by Stirling's formula. We shall give here only the results of successive simplifications. At each step we shall omit the factors free of  $k$  or  $w$ , since their product may be reconstructed from the final exponential form. Thus instead of the right member of (7) we can consider the expression:

$$(16) \quad m^{-1}C_{k-1} \cdot n^{-1}C_{k-1}.$$

Omitting factors free of  $k$ , we get

$$(17) \quad \frac{1}{(k-1)!(m-k)!(k-1)!(n-k)!}$$

and by Stirling's formula, since  $k$  and  $m$  are both large:

$$(18) \quad \frac{1}{k^{k(2k'+1)}(m'-k')^{(m'-k'+\frac{1}{2})}(n'-k')^{(n'-k'+\frac{1}{2})}}.$$

Now apply (14). We obtain

$$(19) \quad \left(\sqrt{m'}w + \frac{m'}{1+\alpha'}\right)^{-2\sqrt{m'}w - \frac{2m'}{1+\alpha'} - 1} \cdot \left(-\sqrt{m'}w + \frac{m'\alpha'}{1+\alpha'}\right)^{\sqrt{m'}w - \frac{m'\alpha'}{1+\alpha'} - \frac{1}{2}} \\ \cdot \left(-\sqrt{m'}w + \frac{m'}{\alpha'(1+\alpha')}\right)^{\sqrt{m'}w - \frac{m'}{\alpha'(1+\alpha')} - \frac{1}{2}}.$$

Dividing inside the parentheses by  $\frac{m'}{1+\alpha'}$ ,  $\frac{m'\alpha'}{1+\alpha'}$ ,  $\frac{m'}{\alpha'(1+\alpha')}$ , respectively, and again omitting factors free of  $w$ , we get

$$(20) \quad \left(1 + \frac{(1+\alpha')w}{\sqrt{m'}}\right)^{-2\sqrt{m'}w - \frac{2m'}{1+\alpha'} - 1} \cdot \left(1 - \frac{(1+\alpha')w}{\alpha'\sqrt{m'}}\right)^{\sqrt{m'}w - \frac{m'\alpha'}{1+\alpha'} - \frac{1}{2}} \\ \cdot \left(1 - \frac{\alpha'(1+\alpha')w}{\sqrt{m'}}\right)^{\sqrt{m'}w - \frac{m'}{\alpha'(1+\alpha')} - \frac{1}{2}}.$$

Taking logarithms, expanding in powers of  $\frac{w}{\sqrt{m'}}$  and neglecting terms in  $\frac{w^3}{m'^{\frac{3}{2}}}$  and higher orders, the results are

$$(21) \quad -\left(2\sqrt{m'}w + \frac{2m'}{1+\alpha'} + 1\right)\left(\frac{(1+\alpha')w}{\sqrt{m'}} - \frac{(1+\alpha')^2w^2}{2m'}\right) \\ -\left(\sqrt{m'}w - \frac{m'\alpha'}{1+\alpha'} - \frac{1}{2}\right)\left(\frac{(1+\alpha')w}{\alpha'\sqrt{m'}} + \frac{(1+\alpha')^2w^2}{2\alpha'^2m'}\right) \\ -\left(\sqrt{m'}w - \frac{m'}{\alpha'(1+\alpha')} - \frac{1}{2}\right)\left(\frac{\alpha'(1+\alpha')w}{\sqrt{m'}} + \frac{\alpha'^2(1+\alpha')^2w^2}{2m'}\right)$$

which equals

$$(22) \quad -\frac{w^2(1+\alpha')^3}{2\alpha'} + O(m'^{-\frac{1}{2}}).$$

The proof of the fact that the distribution of  $w$  converges uniformly to the normal distribution with zero mean and variance  $\frac{\alpha'}{(1+\alpha')^3}$  can be carried out in the same way as the classical proof that the binomial law converges to the normal distribution.

It is obvious that

$$w^* = \frac{k - \frac{m}{1 + \alpha}}{\sqrt{m}}$$

has the same distribution as  $w$ . From this and from the fact that  $U = 2k$  or  $2k - 1$  THEOREM I follows.

In using conventional tables of the Gaussian function to make tests of significance on  $U$  when  $m$  and  $n$  are large, the reader is urged not to forget that the critical region of  $U$  lies in only one tail of the curve.

**6. An example.** We give here a simple example illustrating the use of the statistic  $U$  and THEOREM I.

Suppose 50 observations were made on  $X$  and 50 observations on  $Y$ . Suppose further that these observations are arranged in ascending order and that the  $i^{\text{th}}$  element of this sequence is said to have the rank  $i$ . The observations on  $X$  occupy the following ranks: 1, 5, 6, 7, 12, 13, 14, 15, 16, 17, 19, 20, 21, 25, 26, 27, 28, 31, 32, 38, 42, 43, 44, 45, 50, 51, 52, 53, 54, 56, 57, 58, 62, 63, 64, 65, 68, 69, 75, 79, 80, 81, 86, 87, 89, 90, 91, 93, 94, 95.

The observations on  $Y$  occupy the remaining ranks.

In this case,  $U = 34$ .

For  $m = n = 50$ ,

$$E(U) = 51,$$

$$\sigma^2(U) = 24.747.$$

The probability of getting 34 runs or less when the distribution functions of  $X$  and  $Y$  are continuous and identical is therefore less than  $5 \cdot 10^{-4}$ .

**7. Consistency.** We shall say that a test is "consistent" if the probability of rejecting the null hypothesis when it is false (i.e., the complement of the probability of a type II error, cf. Neyman and Pearson, [3]) approaches one as the sample number approaches infinity. In the literature of statistics a function of the observations which converges stochastically to a population parameter as the sample number approaches infinity, is called a "consistent" statistic. If a test of a hypothesis about a population parameter is made by a proper use of a consistent (statistic) estimate of the parameter, the test will be consistent also according to our definition, which thus furnishes an extension of the idea of consistency to the case where the alternatives to the null hypothesis cannot be specified by a finite number of parameters.

It is obvious that consistency ought to be a minimal requirement of any good test. It is the purpose of this section to prove that, subject to some slight and from the practical statistical point of view, unimportant, restrictions on the distribution functions, the test furnished by the statistic  $U$  is consistent.

We shall say that the distribution functions  $f(x)$  and  $g(x)$  satisfy the condition A, if, for any arbitrarily small positive  $\delta$ , there exist a finite number of



closed intervals, such that the probability of the sum  $I$  of these intervals is  $> 1 - \delta$  according to at least one of the distribution functions  $f(x)$  and  $g(x)$ , and such that  $f(x)$  and  $g(x)$  have positive continuous derivatives  $f'(x)$  and  $g'(x)$  in  $I$ .

In all that follows, although  $m$  and  $n$  are considered as variables, their ratio  $m/n$  is to be a constant, denoted by  $\alpha$ . Let  $\beta > 0$  denote the level of significance on which the test is to be made, so that, if  $f(x) \equiv g(x)$ ,

$$(23) \quad P\{U < u_0(m)\} = \beta$$

where the critical region for two samples of size  $m$  and  $n$ , respectively, is given by

$$U < u_0(m).$$

THEOREM II. If  $f(x)$  and  $g(x)$  satisfy condition A, and if

$$(24) \quad f(x) \not\equiv g(x),$$

then

$$(25) \quad \lim_{m \rightarrow \infty} P\{U < u_0(m)\} = 1.$$

The proof of this theorem will be given in several stages.

Let  $E\left(\frac{U}{m}; f; g\right)$  and  $\sigma^2\left(\frac{U}{m}; f; g\right)$  denote the mean and variance, respectively, of  $\frac{U}{m}$ , when  $X$  and  $Y$  have the distribution functions  $f(x)$  and  $g(x)$ , respectively, and the sample numbers are  $m$  and  $n$ . Let the set  $x_1 \dots x_m; y_1 \dots y_n$  be arranged in ascending order of magnitude, thus:

$$(26) \quad Z = z_1, z_2, \dots, z_{m+n},$$

where  $z_1 < z_2 < \dots < z_{m+n}$ . The sequence

$$(27) \quad V = v_1, v_2, \dots, v_{m+n}$$

is defined as follows:  $v_i = 0$  if  $z_i$  is a member of the set  $x_1 \dots x_m$  and  $v_i = 1$  if  $z_i$  is a member of the set  $y_1 \dots y_n$ .

LEMMA 1. If the following are fulfilled:

$$\begin{aligned} \text{a)} \quad & f(x) \equiv 0 & x < 0, \\ & f(x) \equiv x & 0 \leq x \leq 1, \\ & f(x) \equiv 1 & x > 1. \\ \text{b)} \quad & g(x) \equiv 0 & x \leq 0, \\ & g(x) \equiv 1 & x \geq 1. \end{aligned}$$

c) The derivative  $g'(x)$  of  $g(x)$  exists, is continuous and positive everywhere in the interval  $0 \leq x \leq 1$ .

d)  $k$  is an arbitrary but fixed positive integer. For every  $m$ ,  $i_{1m} < i_{2m} < \dots < i_{km}$  are a set of  $k$  positive integers subject only to the restriction that the least upper bound  $\gamma$  of the sequence  $\frac{i_{km}}{m+n}$  is less than 1.

Then the expected value

$$E\left(\prod_{j=1}^k v_{i_{jm}}\right) \text{ of } \prod_{j=1}^k v_{i_{jm}}$$

satisfies the inequality

$$(28) \quad E\left(\prod_{j=1}^k v_{i_{jm}}\right) - \prod_{j=1}^k \frac{g'(a_{\lambda_{jm}})}{\alpha + g'(a_{\lambda_{jm}})} < \varphi(m)$$

where  $\lambda_{jm} = \frac{i_{jm}}{m+n}$  and  $a_{\lambda_{jm}}$  ( $j = 1 \dots k$ ) is the root of

$$(29) \quad ma_{\lambda_{jm}} + ng(a_{\lambda_{jm}}) = \lambda_{jm}(m+n)$$

and  $\varphi(m)$  depends only on  $m$  and is such that

$$(30) \quad \lim_{m \rightarrow \infty} \varphi(m) = 0.$$

It is easy to verify that the root  $a_{\lambda_{jm}}$  of (29) exists and is unique.

PROOF: It will be sufficient to show that, for any specified set of values of

$$v_{i_{1m}} \dots v_{i_{(r-1)m}}, \quad v_{i_{(r+1)m}} \dots v_{i_{km}} \quad (r = 1 \dots k)$$

the conditional probability  $P\{v_{i_{rm}} = 1\}$  of the relation in braces satisfies the inequality

$$(31) \quad \left| \frac{g'(a_{\lambda_{rm}})}{\alpha + g'(a_{\lambda_{rm}})} - P\{v_{i_{rm}} = 1\} \right| < \psi(m),$$

where  $\psi(m)$  depends only on  $m$  and is such that

$$(32) \quad \lim_{m \rightarrow 0} \psi(m) = 0.$$

For each  $m$  let

$$(33) \quad V'_m = v'_{i_{1m}}, v'_{i_{2m}} \dots v'_{i_{(r-1)m}}, v'_{i_{(r+1)m}} \dots v'_{i_{km}}$$

be a fixed sequence whose elements are either 0 or 1. We shall consider the conditional probability  $P\{v_{i_{rm}} = s\}$ , ( $s = 0, 1$ ) of the relation in braces subject to the condition that

$$(34) \quad v_{i_{jm}} = v'_{i_{jm}}, \quad (j = 1, 2, \dots (r-1), (r+1), (r+2), \dots k).$$

Let  $a$  and  $b$  be two numbers such that  $0 < a < b < 1$ , and let  $m^*$  be a non-negative integer such that  $m^* \leq m$ , and  $m^* \leq [\gamma(m+n)]$  where  $[\gamma(m+n)]$  denotes the largest integer  $\leq \gamma(m+n)$ . Let  $Q_m(a, b, m^*)$  denote the proba-

bility that, if  $m^*$  observations are made on  $X$  and  $[\gamma(m+n)] - m^*$  observations are made on  $Y$ , the following conditions will be fulfilled:

- (a) the total number of observations  $< a$  is exactly  $i_{rm} - 1$
- (b) all observations are  $< b$
- (c) if the  $[\gamma(m+n)]$  observations are arranged in ascending order and if  $v_j^* = 0$  or 1 according as the  $j^{\text{th}}$  element is an observation on  $X$  or on  $Y$ , then

$$(35) \quad v_{i_{rm}}^* = v'_{i_{rm}} \quad (j = 1, 2, \dots, r-1),$$

and

$$(36) \quad v_{i_{jm}-1}^* = v'_{i_{jm}} \quad (j = r+1, r+2 \dots k).$$

It is easy to see that the probability  $P_0$  of the simultaneous fulfillment of the relations (34) and of  $v_{i_{rm}} = 0$  is given by

$$(37) \quad P_0 = \int_0^1 \int_0^b \sum_{m^*} R_m(a, b, m^*) m' (1-b)^{m'-1} (1-g(b))^{n'} da db,$$

where

$$(38) \quad R_m(a, b, m^*) = {}^m C_{m^*} {}^n C_{[\gamma(m+n)]-m^*} \frac{dQ_m}{db}(a, b, m^*),$$

$$(39) \quad m' = m - m^*,$$

and

$$(40) \quad n' = n - [\gamma(m+n)] + m^*.$$

Similarly, the probability  $P_1$  of the simultaneous fulfillment of the relations (34) and of  $v_{i_{rm}} = 1$  is given by

$$(41) \quad P_1 = \int_0^1 \int_0^b \sum_{m^*} R_m(a, b, m^*) n' g'(a) (1-b)^{m'} (1-g(b))^{n'-1} da db.$$

Then

$$(42) \quad \frac{P\{v_{i_{rm}} = 0\}}{P\{v_{i_{rm}} = 1\}} = \frac{P_0}{P_1}.$$

Let  $n_0 = \sum_{j > [\gamma(m+n)]} v_j$  and  $m_0 = m + n - [\gamma(m+n)] - n_0$ . The variables  $(z_{i_{rm}} - a_{\lambda_{rm}}), (z_{[\gamma(m+n)]} - a_\gamma), \left(\frac{m_0}{n_0} - \frac{\alpha(1-a_\gamma)}{(1-g(a_\gamma))}\right)$  all converge stochastically to zero.

Let  $P_0(\epsilon)$  and  $P_1(\epsilon)$  denote the values of the right members of (37) and (41), respectively, if the integration is restricted to the region where  $a \leq b$ ,  $|a - a_{\lambda_{rm}}| < \epsilon$ ,  $|b - a_\gamma| < \epsilon$  and the summation is restricted to those values

of  $m^*$  for which  $\left| \frac{m'}{n'} - \frac{\alpha(1 - a_\gamma)}{(1 - g(a_\gamma))} \right| < \epsilon$ . Hence, because of the aforementioned stochastic convergence, for all sufficiently large  $m$

$$(43) \quad |P_s(\epsilon) - P_s| < \epsilon \quad s = 1, 2.$$

Since  $P_s > 0$ , for sufficiently large  $m$ , also

$$(44) \quad \frac{P_0(\epsilon)}{P_1(\epsilon)} - \frac{P_0}{P_1} < \epsilon.$$

Since  $g(x)$  and  $g'(x)$  are continuous in the interval  $[0, 1]$  and hence uniformly continuous, it is clear that

$$(45) \quad \left| \frac{P_0(\epsilon)}{P_1(\epsilon)} - \frac{g'(a_{\lambda, m})}{g'(a_{\lambda, r, m})} \right| < c\epsilon,$$

where  $c$  is a fixed constant independent of  $m$ . From (44) and (45) it follows easily that, for any arbitrarily small  $\epsilon'$ ,

$$(46) \quad \left| \frac{P_0}{P_1} - \frac{\alpha}{g'(a_{\lambda, r, m})} \right| < \epsilon'$$

for sufficiently large  $m$ .

Since  $P\{v_{i, r, m} = 1\} = \frac{P_1}{P_0 + P_1}$ , the required relation (31) follows. This completes the proof of LEMMA 1.

LEMMA 2. *If conditions a, b, and c of Lemma 1 are satisfied, then*

$$(47) \quad \lim_{m \rightarrow \infty} E\left(\frac{U}{m}; f; g\right) = 2 \int_0^1 \frac{g'(x)}{\alpha + g'(x)} dx$$

and

$$(48) \quad \lim_{m \rightarrow \infty} \sigma^2\left(\frac{U}{m}; f; g\right) = 0.$$

PROOF: Since

$$(49) \quad \begin{aligned} \frac{U}{m} &= \frac{1}{m} + \frac{1}{m} \sum_{j=2}^{m+n} (v_j - v_{j-1})^2 \\ &= \frac{1 + v_1 + v_{m+n}}{m} + \frac{2}{m} \sum_{j=2}^{m+n-1} v_j - \frac{2}{m} \sum_{j=2}^{m+n} v_{j-1} v_j, \end{aligned}$$

we have from LEMMA 1,

$$(50) \quad \begin{aligned} E\left(\frac{U}{m}\right) &= \frac{2}{m} \left[ \sum_i \frac{g'(a_{jm})}{\alpha + g'(a_{jm})} - \sum_i \left( \frac{g'(a_{jm})}{\alpha + g'(a_{jm})} \right)^2 \right] + \eta(m) + \eta^*(\gamma) \\ &\quad \frac{2}{m} \sum \left[ \frac{\alpha g'(a_{jm})}{(\alpha + g'(a_{jm}))^2} \right] + \eta(m) + \eta^*(\gamma), \end{aligned}$$

where

$$(51) \quad \lim_{m \rightarrow \infty} \eta(m) = \lim_{\gamma \rightarrow 1} \eta^*(\gamma) = 0$$

and  $a_{jm}$  is the root of the equation

$$(52) \quad ma_{jm} + n\gamma(a_{jm}) = j \quad (j = 2 \dots m + n).$$

From equation (52) it follows that

$$(53) \quad \lim_{m \rightarrow \infty} (a_{jm} - a_{(j-1)m})(m + n\gamma'(a_{jm})) = 1$$

uniformly in  $j$ . Since  $\gamma$  may be chosen arbitrarily near to 1, the required result (47) follows easily from (50).

It remains to consider the variance of  $\frac{U}{m}$ . The expression

$$\frac{1 + v_1 + v_{m+n}}{m} + \frac{2}{m} \sum_{j=2}^{m+n-1} v_j$$

differs from  $\frac{2}{\alpha}$  by at most  $\frac{1}{m}$ , so that its variance converges to zero with  $m \rightarrow \infty$ .

In order to prove (48), it will be sufficient to show that the variance of

$$(54) \quad W = \frac{1}{m} \sum_{j=2}^{m+n} v_{j-1} v_j$$

goes to zero with increasing  $m$ . From LEMMA 1 it follows that

$$(55) \quad -z(m) < [E(v_i v_j v_k v_l) - E(v_i v_j)E(v_k v_l)] < z(m),$$

where  $\lim_{m \rightarrow \infty} |z(m)| = 0$ , provided only that the integers  $i, j, k, l$  are distinct and  $< \gamma(m + n)$ . The variance of  $mW$  is the sum of terms of the type occurring in (55). The number of terms for which  $i, j, k, l$  are distinct is of the order  $m^2$ . All other terms are of size at most 2 and their number is of the order  $m$ . Since the number  $\gamma$  may be chosen arbitrarily near to 1, the variance of  $W$  converges to zero with  $m \rightarrow \infty$ .

This proves LEMMA 2.

LEMMA 3. *If conditions a, b, and c of Lemma 1 are fulfilled, and if (24) holds, then*

$$(56) \quad T = \int_0^1 \frac{g'(x)}{\alpha + g'(x)} dx < \frac{1}{1 + \alpha}.$$

Let  $a_1 < a_2$  be any two real numbers and designate  $\frac{a_1 + a_2}{2}$  by  $a_2$ . Let  $F(x)$  be defined as follows:

$$(57) \quad \begin{aligned} F(a_1) &= 0, \\ F(x) &= (x - a_i)b_i + F(a_i), \end{aligned} \quad (a_i \leq x \leq a_{i+1}; i = 1, 2).$$

Let  $c$  be defined by

$$(58) \quad F(a_3) = c(a_3 - a_1).$$

Then it is easy to verify that the maximum of

$$(59) \quad T^* = \int_{a_1}^{a_3} \frac{F'(x)}{\alpha + F'(x)} dx$$

with respect to  $b_1$  and  $b_2$ , subject to the restrictions that  $b_1$  and  $b_2$  be non-negative, and that  $a_1$ ,  $a_3$  and  $c$  be fixed ( $c > 0$ ), occurs when and only when

$$(60) \quad b_1 = b_2 = c.$$

Now define

$$(61) \quad \begin{aligned} P_{ij} &= \frac{i}{2^j}, & P_{0j} &= 0, \\ l_{ij} &= \frac{g(P_{ij}) - g(P_{(i-1)j})}{2^j} \end{aligned}$$

and

$$S_i = \frac{1}{2^j} \sum_{i=1}^{2^j} \frac{l_{ij}}{\alpha + l_{ij}}, \quad (i = 1, 2, \dots, 2^j; j = 0, 1, 2, \dots).$$

Repeated application of the result of the previous paragraph easily gives

$$(62) \quad S_i \geq S_{i+1}.$$

From (24) it follows that there exists a positive integer  $j'$  such that  $S_{j'} > S_{j'+1}$ . Obviously

$$(63) \quad S_0 = \frac{1}{1 + \alpha}$$

and

$$(64) \quad \lim_{j \rightarrow \infty} S_j = T.$$

Hence LEMMA 3 is proved.

*Proof of Theorem II:* Let  $\delta_1 > \delta_2 > \dots > \delta_j > \dots$  be an arbitrary but fixed sequence such that  $\lim \delta_j = 0$ . For  $\delta = \delta_j$ , let  $I_1, \dots, I_{k(j)}$  be a set of closed intervals such that no two intervals have an interior point in common and within which, by condition (A),  $f'(x)$  and  $g'(x)$  exist, are positive, and continuous. Let  $I_{0j}$  be the complementary set (with respect to the whole line). (It is easy to see that, if condition (A) is fulfilled, such a system can be constructed.) Let  $U_i$  ( $i = 1, 2, \dots, k(j)$ ) and  $U_{0j}$  denote, respectively, the runs caused by the observations which fall in the intervals  $I_i$ ,  $I_{0j}$ . Then

$$(65) \quad \left| U - \sum_{i=1}^{k(j)} U_i - U_{0j} \right| \leq 2(k(j)).$$

From condition (A) it follows that, with a probability arbitrarily close to 1, for sufficiently large  $m$ ,

$$(66) \quad U_{0j} < 3pm\delta_j,$$

where 
$$p = \max \left[ 1, \frac{1}{\alpha} \right], \quad (j = 1, 2 \dots).$$

Let  $[a_i \leq x < b_i]$ ,  $i = 1, 2 \dots$  denote the interval  $I_i$ , and let  $m_i$  and  $n_i$  denote the number of observations on  $X$  and  $Y$ , respectively, which fall in the interval  $I_i$ . Then  $\frac{m_i}{m}$  and  $\frac{n_i}{n}$  converge stochastically with increasing  $m$  to  $[f(b_i) - f(a_i)]$  and  $[g(b_i) - g(a_i)]$ , respectively.

Within the interval  $I_i$  ( $i = 1, 2 \dots k$ ) we perform the transformation

$$(67) \quad X^* = f(X), \quad Y^* = f(Y),$$

which leaves  $U_i$  invariant. For fixed  $m_i$ ,  $n_i$  the relative distribution of  $X^*$  is uniform and the relative distribution of  $Y^*$  fulfills condition (c) of LEMMA 1.

Hence from LEMMA 2 we obtain that  $\frac{U_i}{m}$  converges stochastically to

$$(68) \quad \lim_{m \rightarrow \infty} E \left( \frac{U_i}{m}; f; g \right) \leq \frac{2[f(b_i) - f(a_i)][g(b_i) - g(a_i)]}{[g(b_i) - g(a_i)] + \alpha[f(b_i) - f(a_i)]}.$$

It can be verified that the sum of the second members in (68) over all values  $i$  is less than or equal to  $\frac{2}{1 + \alpha}$ .

From (24) and condition (A) we get that, for sufficiently small  $\delta_j$ , there exists at least one interval for which the first member of (68) is less than the second member. Hence

$$(69) \quad \Sigma < \frac{2}{1 + \alpha},$$

where

$$(70) \quad \Sigma = \sum_{i=1}^{\infty} \lim_{m \rightarrow \infty} E \left( \frac{U_i}{m}; f; g \right).$$

Now take  $j$  so large that

$$(71) \quad 3p\delta_j < \epsilon,$$

where

$$(72) \quad 0 < 3\epsilon < \frac{2}{1 + \alpha} - \Sigma.$$

Since  $\frac{U_i}{m}$  converges stochastically to its expected value, from (65), (66), (70), (71), and (72), it follows that, with a probability arbitrarily close to 1, for sufficiently large  $m$ ,

$$(73) \quad \frac{U}{m} < \frac{2}{1+\alpha} - \epsilon.$$

From (23) and THEOREM I we get

$$(74) \quad \lim_{m \rightarrow \infty} \frac{u_0(m)}{m} = \frac{2}{1+\alpha}.$$

THEOREM II follows easily from (73) and (74).

**8. Remarks on a proposed test.** We have already remarked in Section 3 that the test proposed by W. R. Thompson is not consistent. To show this, we shall give two distribution functions  $f(x)$  and  $g(x)$  such that, although these functions will be very different, the probability of rejecting the hypothesis that they are the same will not approach one as the sample number approaches infinity.

Suppose, to simplify the notation, that the observations have been ordered according to size, i.e., that  $x_1 < x_2 < \dots < x_m$  and  $y_1 < y_2 < \dots < y_n$ . Suppose further that  $m = n$ , and that the test is to be made on a level of significance  $\beta > 0$ . In the right member of (2) we need not exhibit  $n$  and shall replace  $k$  and  $k'$  by  $k(m)$  and  $k'(m)$  to show the dependence on  $m$ . We have, under the null hypothesis,

$$(75) \quad P\{x_{k(m)} < y_{k'(m)}\} = \psi(m, k(m), k'(m)) = \beta.$$

The sequence  $\frac{k(m)}{m}$  is bounded, so that there exists a monotonically increasing subsequence  $m_1, m_2, \dots$  of the sequence of integers  $1, 2, \dots$  and a number  $h$ ,  $0 \leq h \leq 1$ , such that

$$(76) \quad \lim_{i \rightarrow \infty} \frac{k(m_i)}{m_i} = h.$$

It is easy to see that then also

$$(77) \quad \lim_{i \rightarrow \infty} \frac{k'(m_i)}{m_i} = h.$$

We shall now assume that  $0 < h < 1$ . If  $h = 0$  or  $1$  only a trivial alteration will be needed in the argument to follow. Let  $\epsilon$  and  $\delta$  be arbitrarily small positive numbers. We now consider two populations, A and B described as follows:

$$A) \quad f(x) \equiv g(x) \equiv x \quad (0 \leq x \leq 1),$$

$$B) \quad f(x) = x \quad (0 \leq x \leq 1),$$

$$g(x) \equiv g(a_i) + \frac{(x - a_i)(g(a_{i+1}) - g(a_i))}{(a_{i+1} - a_i)} \quad (a_i \leq x \leq a_{i+1}; i = 0, 1, \dots, 4),$$



where

$$\begin{array}{ll}
 a_0 = 0 & g(a_0) = 0 \\
 a_1 = h - 2\delta > 0 & g(a_1) = 0 \\
 a_2 = h - \delta & g(a_2) = a_2 \\
 a_3 = h + \delta < 1 - \delta & g(a_3) = a_3 \\
 a_4 = 1 - \delta & g(a_4) = a_3 \\
 a_5 = 1 & g(a_5) = 1
 \end{array}$$

The definition of  $f(x)$  and  $g(x)$  outside the interval  $0 \leq x \leq 1$  is obvious. It will be shown that even for such different populations as A and B and for samples of size greater than that of any arbitrarily assigned number, the probability of rejecting the null hypothesis if B is true will be at most  $\beta + \epsilon$ .

Let  $h_1, h_2, h_3$  denote the number of observations on  $X$  which fall in the intervals  $0 < x \leq a_2, a_2 < x \leq a_3, a_3 < x \leq 1$ , respectively ( $m$  fixed, of course). Let  $h'_1, h'_2, h'_3$  be the corresponding numbers for  $Y$ . For a fixed  $m$ , the probability of a set  $h_1, h_2, h_3, h'_1, h'_2, h'_3$  is the same whether the sample be drawn from the population A or B. From (76), (77), and multinomial law it follows that for all sufficiently large  $m_i$  the probability is at least  $1 - \epsilon$  of the occurrence of a set  $h_1, h_2, h_3, h'_1, h'_2, h'_3$  for which  $x_{k(m_i)}$  and  $y_{k'(m_i)}$  will both fall in the interval  $a_2 < x \leq a_3$ . Furthermore it is obvious that for all samples with fixed  $h_2, h'_2$  the distribution within the interval  $a_2 < x \leq a_3$  is the same whether the sample came from the population A or B. Hence even when the sample is drawn from the population B, the first member of (75) is  $< \beta + \epsilon$ . This completes the proof of the inconsistency of the test based on (75).

This test is consistent if the alternatives to the null hypothesis are limited, for example, to those where  $g(x) \equiv f(x + c)$ ,  $c$  a constant.

#### REFERENCES

- [1] WILLIAM R. THOMPSON, *Annals of Math. Stat.*, Vol. 9, (1938), p. 281.
- [2] MAURICE FRÉCHET, *Généralités sur les Probabilités. Variables aléatoires*, Paris, (1937).
- [3] J. NEYMAN AND F. S. PEARSON. *Statistical Research Memoirs*. University College, London. Vol. 1, (1936).

COLUMBIA UNIVERSITY,  
NEW YORK, N. Y.

## THE SUBSTITUTIVE MEAN AND CERTAIN SUBCLASSES OF THIS GENERAL MEAN

BY EDWARD L. DODD

**1. Introduction.** No general agreement has been reached, so far as I know, as to what constitutes a mean. A *necessary condition* which appears to meet with general approval is that a single-valued mean of a set of numbers all equal to a constant  $c$  should itself be equal to  $c$ . However, there appears to be some valid objection against imposing any *other* proposed condition as *necessary*.

Of course, intermediacy is a condition that suggests itself at once. Indeed, in certain mean value theorems in general analysis—such as the First Theorem of the Mean for integral calculus, which I mention in Section 3—intermediacy is the main feature.

However, O. Chisini [1] insisted that intermediacy or internality is not the chief characteristic of a statistical mean. Rather, a mean is a number to take the place, by substitution, of each of a set of numbers in general different. Such a mean may well be called a *representative* or *substitutive* mean.

Chisini defined  $m$  to be a mean of  $x_1, x_2, \dots, x_n$ , relative to a function  $F$ , provided that

$$(1.1) \quad F(m, m, \dots, m) = F(x_1, x_2, \dots, x_n).$$

If, for example,

$$(1.2) \quad F(x_1, x_2, \dots, x_n) = \Sigma x_i^2 = \Sigma m^2 = nm^2,$$

the mean  $m$  thus obtained is the root-mean-square

$$(1.3) \quad m = \pm [(1/n)\Sigma x_i^2]^{1/2}.$$

The choice of  $F$ , Chisini noted, depended upon the use to be made of the mean.

Suppose now that  $f(x_1, x_2, \dots, x_n)$  is such a function that one value of

$$(1.4) \quad f(x, x, \dots, x) = x.$$

And suppose that this  $f$  is taken as a particular  $F$  for (1.1) to determine a mean  $m$  *implicitly*; thus

$$(1.5) \quad f(m, m, \dots, m) = f(x_1, x_2, \dots, x_n).$$

Then, from (1.5) and (1.4) it follows that one value of

$$(1.6) \quad f(x_1, x_2, \dots, x_n) = m.$$

And thus  $f$  determines the mean  $m$  both *explicitly* and *implicitly*.

It should be noted that the  $F = \Sigma x_i^2$  in (1.2) is *not* itself a mean of the  $x_i$ .

If, in (1.2), we take  $x_1 = -2$ ,  $x_2 = 1$ ,  $x_3 = 1$ , then the double-valued mean  $m = \pm 2^{1/2}$  results. Now  $-2^{1/2}$  is *internal*; e.i.  $-2 < -2^{1/2} < 1$ ; but  $2^{1/2}$  is *external*, for  $2^{1/2} > 1 > -2$ . But since here  $\Sigma x_i = 0$ , it follows also that the *standard deviation* of  $-2, 1, 1$ , is the *external mean*  $2^{1/2}$ . Chisini [1], indeed, used the root mean square to show the possibility of external means. External means have been noted by other writers, [2-7].

It is noteworthy that a number of writers [8-12] have used the condition (1.4) (in general, with  $f$  single-valued) as one of a set of axioms to characterize particular means. Sometimes, this has appeared in weaker form as  $f(1, 1, \dots, 1) = 1$ .

This paper will be concerned primarily with the mean of a finite number  $n$ , of variates,  $x_1, x_2, \dots, x_n$ . Possible generalizations will be mentioned briefly in Section 8.

In the conception of the substitutive mean,  $m$ , as I have been using it for some time, emphasis is laid upon the *explicit* form for  $m$ ; and provision is made for *multiple* values.

**DEFINITION OF THE SUBSTITUTIVE MEAN.** Let  $f(x_1, x_2, \dots, x_n)$  be a function of  $n$  variables,  $x_1, x_2, \dots, x_n$  defined at least for one set of equal values,  $x_i = k$ . If  $c$  is any number such that  $f(c, c, \dots, c)$  is defined, let one value of

$$(1.7) \quad f(c, c, \dots, c) = c.$$

Then  $f(x_1, x_2, \dots, x_n)$  will be said to be a substitutive mean of  $x_1, x_2, \dots, x_n$ .

If an original formulation of a problem does not assign to a function a value when the variables are all equal, it is sometimes possible to assign such values by continuity considerations, such as are commonly used in the "evaluation" of indeterminate forms. This will be discussed in Section 6.

In the following, when the word mean is used, it will designate the substitutive mean as defined above.

**2. Classification of Means already made.** Some general classes of means have already been distinguished. One important basis for a classification of means is the kind of data to be used. The data may be only qualitatively distinguishable. Then numbers may be assigned to qualities. For dealing in a very general way with all kinds of data, C. Gini and L. Galvani [13], and G. Pietra [14], distinguished between data in rectilinear series, in cyclical series, and in unconnected series. These three classes are associated respectively with the straight line, the circle, and a regular polyhedron (in three dimensions, the regular tetrahedron, and in  $n$  dimensions, a polyhedron with  $n + 1$  vertices each at the same distance from each of the other  $n$  vertices).

For one definition of the arithmetic mean of a cyclical series, Gini uses the center of gravity principle; and this mean is computed with the aid of sines and cosines. By mechanical means, such an arithmetic mean of dates—for example,

of dates of weddings—as days of a year can be found. On the rim of a wheel delicately suspended and marked off for the 365 days or 366 days of a year, let small weights proportional to the number of weddings on a day be placed in the spaces assigned to the individual days. Then when the wheel comes to rest, the arithmetic mean of the dates will be found at the lowest point of the rim. In the special case where the center of gravity of the system is at the center of the circle, the mean is indeterminate, or we may say that every day is a mean day.

Also, for cyclical series the arithmetic mean and the median are defined by other methods, using such principles as minimizing the sum of the squares of deviations or the sum of the absolute deviations.

The properties of means may be made the basis of a classification, either those properties which have been evolved by writers [8–12], [15–18] who have characterized specific means by sets of axioms, or those properties which seem of special importance in making distinctions. Two such properties will now be mentioned.

Gini [19] recognizes two large classes of means: “A) *medie ferme*, B) *medie lasche*,” the latter (loose) class including the median and mode for which values do not depend upon all the data. To describe this latter mean  $m$  of arguments  $x_i$ , we might write  $\partial m / \partial x_i = 0$  as applying to several if not most of the arguments over wide ranges instead of at isolated points.

Subclasses of A or firm means as given by Gini will be discussed in Section 4.

Another rather large classification distinguishes between simple means and their weighted forms. In a case often encountered, where the weights are whole numbers indicating frequencies of occurrence this distinction is of little significance. In the more general case, however, where weights may give ratings of the efficiency of measuring instruments or the weights may be negative [6, 20], more direct attention needs to be paid the weighted forms.

To supplement classifications already proposed, I am indicating in the next section a descent from the substitutive mean, the most general of all means, down through two classes of means less general, which I am calling the summational mean and the quasi-arithmetic mean, to the more specific mean known as the associative mean, studied in particular by M. Nagumo, [21] A. Kolmogoroff, [22] and B. de Finetti, [2].

The foregoing subclasses of the general or substitutive mean are based primarily on structure, the way the mean is formed.

**3. The Summational Mean, Quasi-Arithmetic Mean, and Associative Mean.** The summational mean, now to be defined, is a generalization of the weighted arithmetic mean.

$$(3.1) \quad W = \frac{c_1 x_1 + c_2 x_2 + \dots + c_n x_n}{c_1 + c_2 + \dots + c_n}, \quad \Sigma c_i \neq 0.$$

It is to be noted that although  $W$  is not a symmetric function of  $x_i$ ,  $W$  is a symmetric function of  $c_i x_i$ . In the generalization  $Q$ , the following features of  $W$  are retained:

1. Certain weights  $c_i$  being given,  $Q$  is a symmetric function of  $c_i x_i$ .
2. This  $Q$  may be determined from sums of  $n$  terms, each term involving one and only one  $x_i$ .

DEFINITION. Let  $\Sigma$  denote a summation for  $i = 1, 2, \dots, n$ . Suppose that

$$(3.2) \quad F\{y, \Sigma f_1(c_i x_i, y), \Sigma f_2(c_i x_i, y), \dots, \Sigma f_k(c_i x_i, y)\} = 0$$

has a solution,  $y = Q$  which is a substitutive mean of  $x_1, x_2, \dots, x_n$ . Then  $Q$  will be called a summational mean of  $x_1, x_2, \dots, x_n$ , relative to the functions  $f_1, f_2, \dots, f_k$ , and  $F$ .

Sometimes it is possible to express  $Q$  as

$$(3.3) \quad Q = G\{\Sigma g_1(c_i x_i), \Sigma g_2(c_i x_i), \dots, \Sigma g_k(c_i x_i)\}.$$

Among summational means, those of most frequent use involve in a special way but one summation. Thus with  $\psi(x)$  a function, which would usually be taken as continuous, this  $m$  satisfies

$$(3.4) \quad \psi(m) \Sigma c_i = \Sigma c_i \psi(x_i).$$

But this, with  $c_i > 0$ , is just an algebraic analogue or prologue to the First Theorem of the Mean for integral calculus—the  $c_i$  to be replaced by a positive integrable function. Without further specification, this mean  $m$  may have an uncountably infinite number of values. But if it be required that  $\psi(x)$  be a continuous increasing function, and that  $c_i > 0$ , then  $m$  is unique.

In a series of papers, C. E. Bonferroni [20], [23–27] used means such as  $m$  in (3.4) for statistical and actuarial problems. And, as he had in mind [28] distinctly the notion of substitution, he was in a sense a forerunner of Chisini. E. L. Dodd [29] made use of a mean  $m$  defined with the aid of  $n$  continuous increasing functions  $\psi_i(x)$ , thus:

$$(3.5) \quad \Sigma c_i \psi_i(m) = \Sigma c_i \psi_i(x_i), \quad c_i > 0.$$

If  $g_i(x) = c_i \psi_i(x)$ , this can be written

$$(3.6) \quad \Sigma g_i(m) = \Sigma g_i(x_i).$$

In one paper, C. E. Bonferroni [20], as already noted, used weights which might be either positive or negative.

Some such mean as  $m$  in (3.4) has been used by a number of writers. Here  $\psi(m)$  is a weighted arithmetic mean of  $\psi(x_i)$ ; and thus it is natural to call  $m$  a quasi-arithmetic mean of  $x_i$ .

DEFINITION. Let  $\Sigma c_i \neq 0$ . If  $m$  is a solution of

$$(3.4) \quad \psi(m) \Sigma c_i = \Sigma c_i \psi(x_i),$$

then  $m$  will be called a quasi-arithmetic mean of  $x_i$ , with weights  $c_i$ , and relative to the function  $\psi(x)$ .

Sufficient conditions for the existence of this mean  $m$  are: (1) That  $\psi(x)$  be continuous in the interval  $I$ , finite or infinite, in which the observations  $x_i$  lie; (2) That either  $c_i > 0$  for each  $i$ , or that  $\psi(x)$  take on all real values, as  $x$  runs through  $I$ .

It will be helpful to picture geometrically the double transformation or mirroring represented by (3.4). Points  $x_i$  on the horizontal axis are carried vertically to the curve  $y = \psi(x)$  and then reflected horizontally to the  $y$  axis. For the points  $y_i$ , on the  $y$  axis thus obtained the arithmetic mean  $\bar{y}$  or "center of gravity" is obtained. Then  $\bar{y}$  is carried horizontally to the curve and reflected vertically to the  $x$ -axis. The abscissas  $m$  of points on the  $x$ -axis thus obtained are means of the given  $x_i$ , relative to this  $\psi(x)$ .

It may happen (Dodd [3 p. 746]) that the curve  $y = \psi(x)$  contains horizontal segments, as in the curve for temperature  $y$  of ice-water-steam which has absorbed a quantity  $x$  of heat. In this case the mean  $m$  may be an "interval," an uncountable set of real numbers. Indeterminateness over an interval is a well known feature of the median of an even number of variates. In fact, a paper of D. Jackson [30] was for the purpose of indicating one method of selecting a single value from this interval of indeterminateness, as a median.

It may be noted that a mean of  $n$  variables becomes, when  $n = 1$ , a function of a single variable; and thus it appears possible to implant in a mean of  $n$  variables almost any peculiarity found in a function of one variable.

A special case of the quasi-arithmetic mean is the associative mean  $m$  which under some general conditions has been shown [2, 21, 22] to satisfy

$$(3.7) \quad n\psi(m) = \sum \psi(x_i), \quad i = 1, 2, \dots, n;$$

where  $\psi(x)$  is a continuous increasing function.

If  $f_n(x_1, x_2, \dots, x_n)$  is an associative mean, then by definition,  $f_n(x_1, x_2, \dots, x_n)$  is unaltered when any  $k$  of the  $n$  variates are each replaced by the mean  $f_k$  of that set.

**4. The Gini means as summational.** Having distinguished firm means from loose means, Gini [19] noted that in the former class, a variate might appear as a base, as an exponent, or both as base and exponent. In general, these variates are to be positive. Gini then listed ten means of a decidedly broad character, some of them generalizing the combinatorial means treated by A. Durand [31] and O. Dunkel [32]. See also G. Pietra [37].

These ten means involve only the four simple arithmetic operations and root extraction. For many purposes they are best expressed in the form given by the author. However, to show that these means are summational, logarithms will be used to reduce products to sums.

Let

$$\begin{aligned}
 S^p &= \sum x_i^p & i &= 1, 2, \dots, n; \\
 {}_nC_c &= n!/c!(n-c)!, \text{ a binomial coefficient;} \\
 P_c &\text{ be any one of the } {}_nC_c \text{ products of } c \text{ different elements taken from} \\
 (4.1) \quad & x_1, x_2, \dots, x_n; \\
 P_c^p &= (P_c)^p, \text{ the } p^{\text{th}} \text{ power of } P_c; \\
 Z_c &= \sum P_c, \text{ the sum of all the } {}_nC_c \text{ products } P_c; \\
 Z_c^p &= \sum P_c^p.
 \end{aligned}$$

In the expressions which follow, it is assumed that the denominators are not zero.

The ten means, as defined in Gini's Equations I, II,  $\dots$ , X, will be designated here by  $m_1, m_2, \dots, m_{10}$ ; and their logarithms, with base arbitrary, will now be given.

$$\begin{aligned}
 \log m_1 &= (\log S^p - \log n)/p \\
 \log m_2 &= (\log Z_c - \log {}_nC_c)/c \\
 \log m_3 &= (\log Z_c^p - \log {}_nC_c)/cp \\
 \log m_4 &= (\log S^p - \log S^q)/(p - q) \\
 (4.2) \quad \log m_5 &= \sum x_i^p \log x_i / S^p \\
 \log m_6 &= (\log Z_c - \log Z_d - \log {}_nC_c + \log {}_nC_d)/(c - d) \\
 \log m_7 &= (\log Z_c^p - \log Z_d^p - \log {}_nC_c + \log {}_nC_d)/(c - d)p \\
 \log m_8 &= (\log Z_c^p - \log Z_c^q)/c(p - q) \\
 \log m_9 &= \sum P_c^q \log P_c / cZ_c^q \\
 \log m_{10} &= (\log Z_c^p - \log Z_d^q - \log {}_nC_c + \log {}_nC_d)/(cp - dq).
 \end{aligned}$$

As noted by the author, the foregoing include some well known special means. Thus,  $m_1$  is the power mean, which for  $p = 1, 2, -1$ , becomes respectively the arithmetic mean, the root mean square, and the harmonic mean. If  $p \rightarrow 0$ , then the limit of  $m_5$  and of  $m_7$  is the geometric mean. If  $p = 0, 1, 2$ , and  $q = p - 1$ , then  $m_4$  is respectively the harmonic, the arithmetic, and the contra-harmonic mean.

For each of the ten means, Gini gives an appropriate name. Those involving binomial coefficients are combinatorial, a mean like the contra-harmonic with denominator other than a constant is biplanar, the more simple means monopolar.

When in the following, I show that certain combinatorial expressions may be

replaced by sums, it is not implied that this replacement would simplify computation.

To prove that  $m_1, m_2, \dots, m_{10}$  are all summational means, it may be noted that  $n, p, q, c, d, {}_nC_c$ , and  ${}_nC_d$  are constants. Moreover,  $S^p$  is the symmetric sum of the  $p$ th powers of  $x_i$ , thus with *only one*  $x_i$  in each term, and  $i = 1, 2, \dots, n$ . And, since  $Z_c, Z_c^p, Z_d$ , and  $Z_d^p$  are symmetric polynomials in the  $x_i$ , they may be expressed as polynomials in  $S^1, S^2, \dots$ , by a well known theorem of algebra. Hence among the ten means, the only one that requires special attention is the ninth mean,  $m_9$ .

To show that  $m_9$  is a summational mean, we need only examine the numerator of the right member. Let this numerator be  $N$ .

$$(4.3) \quad N = \Sigma P_c^q \log P_c.$$

Then

$$(4.4) \quad qN = (x_1^q x_2^q \dots x_n^q)(\log x_1^q + \dots + \log x_n^q) + \dots$$

Thus, if we set  $y_i = x_i^q$ , we may write

$$(4.5) \quad qN = (y_1 y_2 \dots y_n)(\log y_1 + \dots + \log y_n) + \dots$$

The coefficient of  $\log y_1$  in this right member is the sum of all products of  $c$  different factors which include  $y_1$ .

Now, let  $Y_r$  be the sum of the products of  $r$  different factors taken from  $y_1, y_2, \dots, y_n$ ; and let  $T_r$  be the sum of the products of  $r$  different factors taken from  $y_2, y_3, \dots, y_n$ . Then it is evident that

$$(4.6) \quad Y_r = T_r + y_1 T_{r-1}; \quad T_r = Y_r - y_1 T_{r-1}.$$

If, now, we set  $Y_0 = 1$ , it follows that

$$(4.7) \quad T_{c-1} = Y_{c-1} - y_1 Y_{c-2} + y_1^2 Y_{c-3} - \dots + (-1)^{c-1} y_1^{c-1} Y_0.$$

Hence, in  $qN$ , the coefficient of  $\log y_1$  is

$$(4.8) \quad y_1 T_{c-1} = y_1 Y_{c-1} - y_1^2 Y_{c-2} + \dots + (-1)^{c-1} y_1^c Y_0.$$

Thus in  $qN$ , the terms containing  $\log y_1$  are

$$(4.9) \quad Y_{c-1} y_1 \log y_1 - Y_{c-2} y_1^2 \log y_1 + \dots + (-1) y_1^c \log y_1.$$

Now let

$$(4.10) \quad U_r = \Sigma y_i^r \log y_i, \quad i = 1, 2, \dots, n.$$

Then,

$$(4.11) \quad qN = Y_{c-1} U_1 - Y_{c-2} U_2 + \dots + (-1)^{c-1} Y_0 U_c.$$

Thus,  $qN$  is here constructed from sums of  $n$  terms with but a *single*  $y_i$  in any term.

Likewise, with  $y_i$  replaced by  $x_i^q$ , a term contains but a single  $x_i$ .



**5. Transformations.** A function  $f(x_1, x_2, \dots, x_n)$  is not in general a mean of its arguments  $x_i$ . However, it is often possible to make a substitution  $x_i = \phi(y_i)$  so that

$$(5.1) \quad f[\phi(y_1), \phi(y_2), \dots, \phi(y_n)] = g(y_1, y_2, \dots, y_n),$$

is a mean of its arguments  $y_i$ .

The required substitution is sometimes obvious, as in the case of the estimate  $s$  of scale

$$(5.2) \quad s = [(1/n)\Sigma(x_i - m)^2]^{1/2} = [(1/n)\Sigma y_i^2]^{1/2}.$$

Here  $s$  is a mean of  $y_i$ , although it is not a mean of  $x_i$ .

**DEFINITION.** Let  $y = \psi(x)$ , in general multiple valued, be defined in an interval  $I$ , finite or infinite, the values of  $y$  lying in an interval  $J$ . Suppose that for each  $y$  in  $J$ , there is at least one  $x$  in  $I$  such that  $\psi(x) = y$ . Let any such  $x$  be designated by  $\phi(y)$ . Then  $\phi(y)$  will be called the inverse of  $\psi(x)$ . It follows that one value of

$$(5.3) \quad \psi[\phi(y)] = y.$$

**THEOREM.** Let

$$(5.4) \quad z = f(x_1, x_2, \dots, x_n),$$

in general multiple valued, be defined when each  $x_i$  is in some interval  $I$ , finite or infinite. With  $x$  in  $I$ , set

$$(5.5) \quad \psi(x) = f(x, x, \dots, x);$$

and suppose that  $y = \psi(x)$  has an inverse,  $x = \phi(y)$  defined in  $J$ . Let  $x_i = \phi(y_i)$  be substituted into  $f$  to form the function

$$(5.6) \quad w = f[\phi(y_1), \phi(y_2), \dots, \phi(y_n)] = g(y_1, y_2, \dots, y_n).$$

Then  $w$  is a mean of  $y_i$ , defined when  $y_i$  is in  $J$ . It is thus a mean of  $\psi(x_i)$ , where  $x_i$  is in  $I$ .

If further,  $\psi(x)$  is a continuous increasing function of  $x$ , then for a given set of  $x_i$ , the values of  $z$  and  $w$  are identical. The same is true for a given set of  $n$  values  $y_i$ .

**PROOF.** If each  $y_i = c$ , a number in  $J$ , then

$$(5.7) \quad f[\phi(y_1), \dots, \phi(y_n)] = f[\phi(c), \dots, \phi(c)] = \psi[\phi(c)].$$

And one value of  $\psi[\phi(c)]$  is  $c$ , from the definition of the inverse function  $\phi(y)$ . Moreover, if a number  $c'$  is taken in  $I$ , then  $\psi(c')$  is some number in  $J$ , which we may call  $c$ ; and the argument above is applicable. Finally, if  $\psi(x)$  is continuous and increasing, then a number  $x_i$  in  $I$  is associated with one and only one  $y_i$  in  $J$ ; and vice versa. Thus  $w$  and  $z$  become identical.

In the foregoing, we started with  $f$  which is not a mean of its arguments  $x_i$ , and obtained  $g$  which is a mean of  $y_i$ . Something like the reverse of this is possible. The last member of (5.2) is a mean of  $y_i$ . It was obtained by treat-

ing  $m$  as a constant, with respect to  $x_i$ . If, however,  $m$  is an estimate for location and is taken as  $(1/n)\Sigma x_i$ , and this is substituted into (5.2) then

$$(5.8) \quad s = \{[(n-1)/n]\Sigma x_i^2 - (2/n)\Sigma x_i x_j\}^{1/2}, \quad i < j.$$

This  $s$  is now not a mean of  $x_i$ ; for if  $x$  equal any constant  $c$ , then  $s = 0$ . Furthermore, there exists no single valued continuous increasing function  $x = \phi(y)$  such that if  $x_i = \phi(y_i)$  is substituted into (5.8),  $s$  will be a mean of the  $y_i$ . Thus the elimination of  $m$  from (5.2) interferes with the status of  $s$  as a mean of the  $x_i$ .

**6. Indeterminate Forms that arise in testing for Means.** Sometimes a function  $f$  is substantially continuous. But the investigation leading to the function fails to assign to the function a value for certain values of the argument  $x$ , or arguments,  $x_1, x_2, \dots, x_n$ . However, values are often assignable which will make the function continuous. This is the usual occurrence when, in curve fitting, parameters are estimated. In general, the measurements are assumed to be not all alike. However, when a general function such as  $\Sigma x_i/n$  for location is obtained, we do not hesitate to assign to this function the value  $c$  when each  $x_i = c$ , to make the function continuous.

As another illustration of "indeterminate forms," consider the Jackson [30] median,  $M$ , of four numbers  $x_1 \leq x_2 < x_3 \leq x_4$ , viz.,

$$(6.1) \quad M = (x_4 x_3 - x_2 x_1)/(x_4 + x_3 - x_2 - x_1).$$

A direct substitution of  $x = c$ , renders  $M$  indeterminate. But if  $x_i \rightarrow c$ , indeed, if merely  $x_2 \rightarrow c$ , and  $x_3 \rightarrow c$ , so also does  $M$ .

In a recent paper, R. Cisbani [33] generalizes means suggested by Dunkel [32] and L. Galvani [34] by setting up

$$(6.2) \quad y_j(x) = \left[ n^{-1} \sum_{i=1}^n (a^i + ih)^{-x/j} \right]^{-1/x}, \quad j \neq 0, \quad x \neq 0;$$

and letting  $n \rightarrow \infty$ . There results an integral with the value

$$(6.3) \quad \bar{y}_j(x) = \left[ \frac{b^{x+j} - a^{x+j}}{(x/j + 1)(b^j - a^j)} \right]^{1/x},$$

for the case,  $x \neq j$ . This mean set up as a mean of an infinite number of variates turns out to be also a mean of the two numbers  $a$  and  $b$ ,—which for  $b = a$  becomes indeterminate. But as  $b$  approaches  $a$ , so also does  $\bar{y}_j(x)$  approach  $a$ . This is also true for the special cases  $x = -j$ , etc.

In testing to see if a function  $m$  of  $x_i$  is a mean of these numbers, a difficulty sometimes arises, because a substitution of  $x_i = c$  and  $m = c$  into the equation which implicitly defines  $m$  will put zeros into denominators. An aid in such testing will now be formulated as a theorem, although the ideas involved are not essentially new.

**THEOREM.** Let  $f(x)$  be a continuous increasing function of  $x$  defined for each real  $x$ . Let

$$(6.4) \quad f(0) = 0.$$

Given  $n$  real distinct numbers

$$(6.5) \quad x_1 < x_2 < \cdots < x_{n-1} < x_n,$$

$n$  positive numbers,  $k_i$ , and a real number  $C$ .

Set

$$(6.6) \quad F(x) = \frac{k_1}{f(x_1 - x)} + \cdots + \frac{k_n}{f(x_n - x)} - C.$$

Then  $F(x) = 0$  has  $n - 1$  real roots  $m_j$ , such that

$$(6.7) \quad x_1 < m_1 < x_2 < m_2 < \cdots < m_{n-1} < x_n;$$

also, a root less than  $x_1$ , provided

$$(6.8) \quad \Sigma k_i / f(+\infty) < C;$$

or a root greater than  $x_n$ , provided

$$(6.9) \quad \Sigma k_i / f(-\infty) > C.$$

**PROOF.** Since  $f(x)$  is a continuous increasing function of  $x$ , so also is  $k_i / f(x_i - x)$ , except for the single value,  $x = x_i$ . So also, then, is  $F(x)$ , except when  $x = x_1$  or  $x_2$  or  $\cdots$  or  $x_n$ . But

$$(6.10) \quad F(x_i + 0) = -\infty; F(x_{i+1} - 0) = +\infty.$$

Hence, between  $x_i$  and  $x_{i+1}$ , there exists a root  $m_i$ , of  $F(x) = 0$ .

Moreover, since

$$(6.11) \quad F(-\infty) = [\Sigma k_i / f(+\infty)] - C; F(x_1 - 0) = +\infty;$$

it follows that there is a root less than  $x_1$ , provided (6.8) is satisfied. Likewise, there is a root greater than  $x_n$  if (6.9) is satisfied.

The use of this theorem in testing for means is simple. Keeping the  $x_i$  distinct, the equation  $F(x) = 0$  determines  $(n - 1)$  numbers,  $m_i$ , such that if  $x_i \rightarrow c$ , so also do these  $m_j \rightarrow c$ . Employing continuity to define  $m_j$  when each  $x_i = c$ , we may say that each  $m_i$  is a mean of  $x_i$ ;  $j = 1, 2, \cdots (n - 1)$ ;  $i = 1, 2, \cdots n$ , when the conditions of this theorem are satisfied. If  $F(x) = 0$  has still another root,  $m$ , this  $m$  will not in general be a mean of  $x_i$ .

**7. Summational Means arising in the Estimation of Parameters of Frequency Distributions.** In curve fitting, the estimation of parameters leads in general to summational means. If the method of moments is used, the first step is to find the moments by summation. I have already considered estimates for location and scale by this method [7], and by the R. A. Fisher method of maxi-

mum likelihood [4]. A further study of the results of the likelihood method will now be made.

By this method, products which first appear are reduced to sums by logarithms, and the means found are, in general, summational. Some idea of the forms of these means can be obtained by examining a rather general form of frequency function which includes the Pearson Type I, and involves parameters with estimates  $p > 0$  and  $q > 0$ , in addition to the location  $m$  and scale  $a$ . Let the observations be  $x_1, x_2, \dots, x_n$ ; let

$$(7.1) \quad t_i = (x_i - m)/a; \quad 0 \leq t \leq 1; \quad a > 0;$$

$$(7.2) \quad y = \frac{1}{a} \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1}(1-t)^{q-1}.$$

The likelihood  $L$  is obtained by multiplying together the  $n$  factors obtained by substituting  $t = t_1, t_2, \dots, t_n$ .

Then

$$(7.3) \quad \begin{aligned} \log L = & -n \log a + n \log \Gamma(p+q) - n \log \Gamma(p) - n \log \Gamma(q) \\ & + (p-1) \sum_1^n \log t_i + (q-1) \sum_1^n \log (1-t_i). \end{aligned}$$

From  $\partial L / \partial m = 0$ , there is obtained

$$(7.4) \quad P \Sigma \frac{1}{x_i - m} + Q \Sigma \frac{1}{x_i - m - a} = 0; \quad P = p - 1, \quad Q = q - 1.$$

Suppose  $P \neq 0$  and  $Q \neq 0$ ; and as a first case, suppose  $P + Q \neq 0$ . If each  $x_i$  is replaced by  $x$ , the above equation leads to  $m = x - (Pa)/(P+Q)$ .

Then  $m$  is a summational mean of

$$(7.5) \quad x'_i = x_i - (Pa)/(P+Q) \quad i = 1, 2, \dots, n;$$

as seen by applying the Theorem in Section 5.

Likewise,  $a$  is a summational mean of

$$(7.6) \quad x''_i = (x_i - m)(P+Q)/P.$$

If  $P \neq 0, Q \neq 0$ ; but  $P+Q=0$ , then (7.4) becomes

$$(7.7) \quad \Sigma \frac{1}{x_i - m - a} = \Sigma \frac{1}{x_i - m}.$$

Now set  $y_i = x_i - m, C = \Sigma 1/y_i$ ; and write (7.7) as

$$(7.8) \quad F(a) = \Sigma \frac{1}{y_i - a} - C = 0.$$

This has the form given in (6.6) with  $x$  replaced by  $a, k_i = 1, f(a) = a$ . If then  $y_1 < y_2 < \dots < y_n$ , there exist  $(n-1)$  solutions  $a_j$  of  $F(a) = 0$  between  $y_1$

and  $y_n$ . And thus keeping the  $y_i$  distinct, if  $y_i \rightarrow c$ , so also do the  $a_i \rightarrow c$ . These  $a_i$  are then means of  $y_i$ , and thus, means of  $x_i - m$ .

In the more general case where  $P + Q \neq 0$ , it is seen also that  $Q$  is a summational mean of

$$(7.9) \quad \left[ \frac{a}{x_i - m} - 1 \right].$$

From  $\partial L / \partial a = 0$ , quite analogous results are obtained. The special case now, however, is given by  $P + Q + 1 = 0 = p + q - 1$ . And, with the continuity interpretation,  $a$  is a mean of  $x_i - m$ ; and moreover,  $m$  is a mean of  $x_i - a$ .

Using now the digamma function

$$(7.10) \quad f(u) = \frac{d}{du} \log \Gamma(u),$$

set

$$(7.11) \quad D(p) = f(p + q) - f(p).$$

The condition  $\partial L / \partial p = 0$ , then leads to

$$(7.12) \quad D(p) = (1/n) \Sigma (-\log t_i), \quad 0 < t_i \leq 1.$$

Now, with  $q > 0$ ,  $D(\infty) = 0$ ,  $D(-1 + 0) = \infty$ ; and  $D(p)$  is a continuous decreasing function of  $p$ , when  $p > -1$ . Then, since  $-\log t_i > 0$ , there is a unique  $p > -1$  to satisfy (6.12).

To be useful, here,  $p$  should be  $> 0$ . But, at all events, the  $p$  thus found is a mean of  $D^{-1}(-\log t_i)$ , where  $D^{-1}$  is inverse to  $D$ .

The digamma function (7.10) appears also in estimating the parameters for the Pearson Type III.

$$(7.13) \quad y = \frac{1}{a} \frac{1}{\Gamma(p+1)} e^{-t} t^p, \quad t = (x - m)/a, \quad p > -1.$$

By setting  $\partial L / \partial p = 0$ , it is found that  $m$  is the *arithmetic* mean of  $x_i - ae^{f(p+1)}$ ;  $a$  is the arithmetic mean of  $(x_i - m)e^{-f(p+1)}$ ; while  $p$  is a summational mean of  $f^{-1}\{\log(x_i - m)/a\} - 1$ , where  $f^{-1}$  is the inverse of  $f$ . From  $\partial L / \partial m = 0$ , it is found that  $m$  is a summational mean of  $x_i - pa$ ;  $a$  is the *harmonic* mean of  $(x_i - m)/p$ ; and  $p$  is the harmonic mean of  $(x_i - m)/a$ . Finally, from  $\partial L / \partial a = 0$ , there is obtained

$$(7.14) \quad (1/n) \Sigma x_i = m + a(p + 1),$$

which makes  $m$ ,  $a$  and  $p$  each an *arithmetic* mean of a simple function of the observations  $x_i$ , when the other two estimates are taken as constants.

Comparison of (5.2) with (5.8) has shown that after complete elimination, estimates may cease to be means. However, it may be noted that  $s$  is more frequently exhibited in the form (5.2) where it is a mean than in the form (5.8) where it is not.

**8. Generalizations.** The extension of results from the discrete or discontinuous case where a mean  $m$  depends upon only a finite number of elements to the continuous case is fairly immediate, with integration taking the place of summation, and a distribution or frequency function taking the place of discrete weights,  $c_i$ . Stieltjes and Lebesgue integrals may be used as well as Riemannian. Such a generalization of the Chisini mean was given by de Finetti [2].

The summational mean, which I have defined as involving possibly several summations, may be generalized likewise.

In terms of set functions, sometimes called functionelles, I gave [35] the following general definition of a mean with a point set  $H$  in mind as a distribution function.

**DEFINITION.** Let  $E$  and  $H$  be sets of numbers. Such a number  $t$  may be a real number or a vector number  $t = (t_1, t_2, \dots, t_k)$ .

Let  $E_t$  be the result of replacing each number of  $E$  by a single number  $t$ .

Then the mean  $m$  of numbers in  $E$ , relative to the set  $H$ , and to a function  $f$ , is given by  $m = f(E, H)$ ; provided that the function  $f$  has been so constructed that for each  $t$  in  $E$ ,  $f(E_t, H) = t$ , or at least one value of this  $f$  is  $t$ . It is to be understood above that when  $E$  is changed to  $E_t$ , the set  $H$  remains unaltered.

This retains the chief feature of  $f(t, t, \dots, t) = t$  in explicit form or of  $f(t, t, \dots, t) = f(t_1, t_2, \dots, t_n)$  in implicit form, where  $t$  is a mean of  $t_1, t_2, \dots, t_n$ .

I used [36] a somewhat less general definition to discuss regression coefficients. All such means may well be called *substitutive* or *representative*.

#### REFERENCES

- [1] O. CHISINI, "Sul concetto di media." *Periodico di Mat.*, Ser. 4, Vol. 9 (1929) pp. 106-116.
- [2] BRUNO DE FINETTI, "Sul concetto di media." *Giornale dell' Inst. Italiano degli Attuari*, Vol. 2 (1931) pp. 369-396. See p. 378.
- [3] E. L. DODD, "The complete independence of certain properties of means." *Annals of Math.*, Vol. 35 (1934) pp. 740-747.
- [4] E. L. DODD, "Internal and external means arising from the scaling of frequency functions." *Annals of Math. Stat.*, Vol. 8 (1937) pp. 12-20.
- [5] E. L. DODD, "Regression coefficients as means of certain ratios." *Am. Math. Monthly*, Vol. 44 (1937) pp. 306-308.
- [6] E. L. DODD, "Index numbers and regression coefficients as means, internal and external." *Report of Third Annual Research Conference on Economics and Statistics*. Cowles Commission for Research in Economics, Colorado Springs, Colo., pp. 13, 14.
- [7] E. L. DODD, "Interior and exterior means obtained by the method of moments." *Annals of Math. Stat.*, Vol. 9 (1938) pp. 153-157.
- [8] O. SUTO, "Law of the arithmetic mean." *Tôhoku Math. Jour.*, Vol. 6 (1914-5) pp. 79-81.
- [9] E. V. HUNTINGTON, "Sets of independent postulates for the arithmetic mean, the geometric mean, the harmonic mean, and the root-mean-square." *Trans. Am. Math. Soc.*, Vol. 29 (1927) pp. 1-22.
- [10] S. NARUMI, "Note on the law of the arithmetic mean." *Tôhoku Math. Jour.*, Vol. 30 (1929) pp. 19-21.
- [11] L. TRODORU, "Sur la définition axiomatique de la moyenne." *Mathematicae*, Vol. 5 (1931) pp. 27-31.
- [12] I. NAKAHARA, "Axioms for the weighted means." *Tôhoku Math. Jour.*, Vol. 41 (1936), pp. 424-434.

- [13] C. GINI, AND L. GALVANI, "Di talune estensioni dei concetti di media ai caratteri qualitativa." *Metron*, Vol. 8 (1929) pp. 3-209.
- [14] G. PIETRA, "The theory of statistical relations with special reference to cyclical series." *Metron*, Vol. 4 (1925) pp. 383-557.
- [15] G. SCHIAPERELLI, "Come si possa giustificare l'uso della media aritmetica nel calcolo dei risultato d'osservazione." *Rendicon to Regio Inst. Lombardo*, Ser. 2, Vol. 40 (1907) pp. 752-764.
- [16] U. BROGGI, "Sur le principe de la moyenne arithmétique." *L'Enseignement*, Vol. 11 (1909) pp. 14-17.
- [17] R. SCHIMMACK, "Der Satz vom arithmetischen Mittel in axiomatischer Begründung." *Math. Annalen*, Vol. 68 (1910) pp. 125-132.
- [18] S. MATSUMURA, "Über die Axiomatik von Mittelbildungen." *Tôhoku Math. Jour.*, Vol. 36 (1933) pp. 260-262.
- [19] C. GINI, "Di una formula comprensiva delle medie." *Metron*, Vol. 13 (1938) pp. 3-22.
- [20] C. E. BONFERRONI, "Sulle medie dedotte da funzioni concave." *Giornale di Mat. Finanziaria*, Vol. 9 (1927) pp. 13-24.
- [21] M. NAGUMO, "Über dine Klasse der Mittelwerte." *Japanese Jour. of Math.*, Vol. 7 (1930) pp. 71-79.
- [22] A. KOLMOGOROFF, "Sur la notion de la moyenne." *Atti Regio Accademia Lincei*, Ser. 6, Vol. 12 (1930) pp. 388-391.
- [23] C. E. BONFERRONI, "Scadenza media e leggi di capitalizzazione." *Giornale di Mat. Finanziaria*, Vol. 3 (1921) pp. 1-25.
- [24] C. E. BONFERRONI, "Sulla scadenza media en senso lato." *Giornale di Mat. Finanziaria*, Vol. 4 (1922) pp. 160-181.
- [25] C. E. BONFERRONI, "La media esponenziale in matematica finanziaria." *Regio Istituto Superiore di Scienze Economiche e Commerciali-Bari*, 1923-24.
- [26] C. E. BONFERRONI, "Della vita matematica come media esponenziale," and Note. *Ibid.*, 1924-25.
- [27] C. E. BONFERRONI, "Sulle medie di potenze." *Giornale degli Economisti e Revista di Statistica*, Vol. 66 (1926) pp. 292-296.
- [28] C. E. BONFERRONI, "A proposito di espressioni generali per le medie." *Giornale degli Economisti e Revista di Statistica*, Vol. 77 (1937) pp. 354-357.
- [29] E. L. DODD, "Functions of measurements under general laws of error." *Skandinavisk Aktuarietidskrift*, Vol. 5 (1922) pp. 133-158.
- [30] DUNHAM JACKSON, "Note on the median of a set of numbers." *Bull. Am. Math. Soc.*, Vol. 27 (1921) pp. 160-164.
- [31] A. DURAND, "Sur un théoreme relatif a des moyennes." *Bull. des Sciences Math.*, Vol. 26 (1902) pp. 181-183.
- [32] O. DUNKEL, "Generalized geometric means and algebraic equations." *Annals of Math.*, Vol. 11 (1909-10) pp. 21-32.
- [33] R. CISBANI, "Contributi alla teoria della medie." *Metron*, Vol. 13 (1938) pp. 23-59.
- [34] L. GALVANI, "Dei limite a cui tendono alcune medie." *Boll. Un. Mat. Italiano*, 1927.
- [35] E. L. DODD, "The chief characteristic of statistical means." Cowles Commission for Research in Economics. *Colorado College Publication*, General Series No. 208 (1936) pp. 89-92.
- [36] E. L. DODD, "Certain coefficients of regression or trend associated with largest likelihood." *Actualités Scientifiques et Industrielles*, No. 740 (1938) pp. 5-14.
- [37] G. PIETRA, "Intorno alle medie combinatorie", *Supplemento statistico, etc.*, Ser. 2, Vol. 4 (1938). This reference is cited in "I Contributi Italiani al Progresso della Statistica," *Societa Italiana per il Progresso delle Scienze*, September 1939 on page 16, as dealing with the expression of power and combinatorial means, monoplane and biplane, as sums of power.

# THE PRODUCT SEMI-INVARIANTS OF THE MEAN AND A CENTRAL MOMENT IN SAMPLES

BY CECIL C. CRAIG

The method developed by the author for calculating the semi-invariants and product semi-invariants of moments in samples from any infinite population<sup>1</sup> is not immediately applicable to the calculation of product semi-invariants of the mean and a central moment in such samples. In the present paper this method is adapted for this purpose so that the calculation of these product semi-invariants becomes routine. As it will be seen, the computing is a little heavier than in the case of central moments alone for results of equal weight. A table of results up to weight ten for the mean and the second, third and fourth central moments is given. The author plans to apply these to a further study of the sampling characteristics of the coefficient of variation and Fisher's  $t$  in samples from non-normal populations.

Let a random sample,  $x_1, x_2, \dots, x_N$  of  $N$  observations be drawn at random from an infinite population characterized by the semi-invariants,  $\lambda_1, \lambda_2, \lambda_3, \dots$ . The sample mean is,

$$\bar{x} = \sum_{i=1}^N x_i / N,$$

and the  $n$ -th central moment of the sample is

$$m_n = \sum_{i=1}^N (x_i - \bar{x})^n / N.$$

Then the product semi-invariants of order  $kl$  of  $x$  and  $m_n$ ,  $S_{kl}(x, m_n)$ , are defined by the formal identity in the parameters  $\vartheta$  and  $\omega$ :

$$(1) \quad (S_{10}\vartheta + S_{01}\omega) + \frac{1}{2!}(S_{10}\vartheta + S_{01}\omega)^{(2)} + \frac{1}{3!}(S_{10}\vartheta + S_{01}\omega)^{(3)} + \dots \equiv \log E(e^{\vartheta x + m_n \omega}),$$

in which  $E$  denotes the mathematical expectation over the set of all such samples and

$$(S_{10}\vartheta + S_{01}\omega)^{(r)} = \sum_{j=1}^r \binom{r}{j} S_{j, r-j}(\bar{x}, m_n) \vartheta^j \omega^{r-j}.$$

<sup>1</sup> "An Application of Thiele's Semi-invariants to the Sampling Problem;" *Metron*, Vol. VII, part IV (1928), pp. 3-75.



If we denote  $E(\bar{x}^k m_n^l)$  by  $M_{ki}$ , we have by definition the further formal identity in  $\vartheta$  and  $\omega$ :

$$E(e^{\bar{x}\vartheta + m_n\omega}) \equiv 1 + (M_{10}\vartheta + M_{01}\omega) + \frac{1}{2!}(M_{10}\vartheta + M_{01}\omega)^{(2)} + \dots$$

in which  $(M_{10}\vartheta + M_{01}\omega)^{(r)}$  is to be expanded in the same manner as  $(S_{10}\vartheta + S_{01}\omega)^{(r)}$  above.

Let us write

$$\delta_i = x_i - \bar{x},$$

and then

$$(2) \quad E(e^{\bar{x}\vartheta + m_n\omega}) = E(e^{(\sum x_i)\vartheta/N + (\sum \delta_i)\omega/N}).$$

(Summations with respect to  $i$  and  $j$  always run from 1 to  $N$ .) Now we define a new set of product semi-invariants,  $\lambda_{rst}\dots$ , of the sum  $\sum x_i$  and the  $N$   $\delta_i$ 's, by means of

$$(\lambda_{10}\vartheta + \sum \lambda_{0i}\omega_i) + \frac{1}{2!}(\lambda_{10}\vartheta + \sum \lambda_{0i}\omega_i)^{(2)} + \dots \equiv \log E(e^{(\sum x_i)\vartheta + \sum \delta_i\omega_i}),$$

in which for example,

$$\begin{aligned} \left(\lambda_{10}\vartheta + \sum_{i=1}^3 \lambda_{0i}\omega_i\right)^{(2)} &= \lambda_{2000}\vartheta^2 + 2\lambda_{1100}\vartheta\omega_1 \\ &\quad + 2\lambda_{1010}\vartheta\omega_2 + \dots + \lambda_{0200}\omega_1^2 + \lambda_{0020}\omega_2^2 + \lambda_{0002}\omega_3^2. \end{aligned}$$

We may set

$$\delta_i = \sum_{j=1}^N a_{ij}x_j \quad \text{with} \quad \begin{cases} a_{ij} = -\frac{1}{N}, & i \neq j \\ a_{ii} = \frac{N-1}{N} \end{cases}$$

Then

$$E(e^{(\sum x_i)\vartheta + \sum \delta_i\omega_i}) = E(e^{\sum x_i(\vartheta + \sum a_{ij}\omega_j)}) = E(e^{\alpha_1 x_1}) \cdot E(e^{\alpha_2 x_2}) \dots E(e^{\alpha_N x_N}),$$

in which

$$\alpha_i = \vartheta + \sum_j a_{ij}\omega_j.$$

It follows then that

$$\begin{aligned} (\lambda_{10}\vartheta + \sum \lambda_{0i}\omega_i) + \frac{1}{2!}(\lambda_{10}\vartheta + \sum \lambda_{0i}\omega_i)^{(2)} \\ + \frac{1}{3!}(\lambda_{10}\vartheta + \sum \lambda_{0i}\omega_i)^{(3)} + \dots \equiv \lambda_1 \Sigma \alpha_i + \lambda_2 \frac{\Sigma \alpha_i^2}{2!} + \lambda_3 \frac{\Sigma \alpha_i^3}{3!} + \dots \end{aligned}$$

from which

$$(\lambda_{10} \vartheta + \sum \lambda_{0i} \omega_i)^{(k+l)} = \lambda_{k+l} \sum_i (\vartheta + \sum_j a_{ij} \omega_j)^{k+l}.$$

From this

$$\begin{aligned} \lambda_{k00\dots 0} &= \lambda_{k0} = N\lambda_k, \\ \lambda_{k10\dots 0} &= \lambda_{k010\dots 0} = \dots = 0, \end{aligned}$$

and generally,<sup>2</sup>

$$(3) \quad \lambda_{kl_1 l_2 \dots l_N} = \frac{\lambda_{k+l}}{N^l} [\sum (-1)^{l-l_i} (N-1)^{l_i}] \quad (l_1 + l_2 + \dots + l_N = l).$$

This is the first result to be used in calculating values of  $S_{kl}$ 's. Note that the value of  $\lambda_{kl_1 l_2 \dots l_N}$  is independent of the order in which a given set of  $l_i$ 's occur.

Calculation of particular  $\lambda_{kl_1 l_2 \dots l_N}$ 's in terms of  $N$  and the semi-invariants of the sampled population is both simple and rapid as one may see from a pair of examples:

$$\lambda_{22} = \lambda_{202} = \lambda_{2002} = \dots$$

(suppressing superfluous zeros in the subscripts)

$$= \frac{\lambda_4}{N^2} [(N-1)^2 + (N-1)] = \frac{N-1}{N} \lambda_4.$$

Then, too,

$$\lambda_{k2} = \frac{N-1}{N} \lambda_{k+2}.$$

For a second example:

$$\begin{aligned} \lambda_{k+3} &= \frac{\lambda_{k+7}}{N^7} [-(N-1)^4 + (N-1)^3 - (N-2)] \\ &= -\frac{(N-2)(N^2 - 3N + 3)}{N^6} \lambda_{k+7}. \end{aligned}$$

Now the semi-invariants,  $S_{kl}$ , can be expressed directly in terms of the product moments,  $\nu_{kl_1 l_2 \dots l_N}$  of the sum  $\sum k_i$  and the  $N\delta$ 's. These product moments are given by the appropriate moment generating function:

$$E(e^{(\sum x_i)\vartheta + \sum \delta_i \omega_i}) = 1 + (\nu_{10}\vartheta + \sum \nu_{0i}\omega_i) + \frac{1}{2!} (\nu_{10}\vartheta + \sum \nu_{0i}\omega_i)^{(2)} + \dots$$

<sup>2</sup> As written this result is valid if at least one of the  $l_i$ 's is zero which is always the case if  $N$ , the size of the sample, is greater than  $l$ . (Cf. the author's paper cited above, p. 17.)

Then it is seen that,

$$E(e^{(\sum x_i)\vartheta + (\sum \delta_i^2)\omega}) = 1 + [\nu_{10}\vartheta + (\sum \nu_{0,ni})\omega] + \frac{1}{2!} [\nu_{10}\vartheta + (\sum \nu_{0,ni})\omega]^{(2)} + \dots,$$

in which

$$[\nu_{10}\vartheta + (\sum \nu_{0,ni})\omega]^{(2)} = \nu_{20}\vartheta^2 + 2(\nu_{1n} + \nu_{10n} + \nu_{100n} + \dots)\vartheta\omega + (\nu_{0,2n} + \nu_{00,2n} + \nu_{000,2n} + \dots)\omega^2$$

etc. and by comparison with (1) and (2), we have

$$(S_{10}\vartheta + S_{01}\omega) + \frac{1}{2!} (S_{10}\vartheta + S_{01}\omega)^{(2)} + \dots \\ \equiv \log \left\{ 1 + \frac{1}{N} [\nu_{10}\vartheta + (\sum \nu_{0,ni})\omega] + \frac{1}{2!N^2} [\nu_{10}\vartheta + (\sum \nu_{0,ni})\omega]^{(2)} + \dots \right\}$$

From this

$$(S_{10}\vartheta + S_{01}\omega)^{(k+l)} \\ (4) \quad \equiv \frac{1}{N^{k+l}} \sum \frac{(-1)^{p-1} (p-1)! (k+l)! [\nu_{10}\vartheta + (\sum \nu_{0,ni})\omega]^r \{ [\nu_{10}\vartheta + (\sum \nu_{0,ni})\omega]^{(2)} \}^s \dots}{(1!)^r (2!)^s \dots r! s! \dots}$$

in which

$$r + s + t + \dots = p,$$

the summation extending over all partitions  $(1'2'3' \dots)$  of  $k + l$ . This, of course, is only the usual formula for semi-invariants in terms of moments appropriately modified. In particular,

$$(S_{10}\vartheta + S_{01}\omega)^{(2)} = \frac{1}{N^2} \{ [\nu_{10}\vartheta + (\sum \nu_{0,ni})\omega]^{(2)} - [\nu_{10}\vartheta + (\sum \nu_{0,ni})\omega]^2 \}.$$

If we write

$$[\nu_{10}\vartheta + (\sum \nu_{0,ni})\omega] = W$$

$$(5) \quad (S_{10}\vartheta + S_{01}\omega)^{(3)} = \frac{1}{N^3} (W^{(3)} - 3W^{(2)}W + 2W^3)$$

$$(S_{10}\vartheta + S_{01}\omega)^{(4)} = \frac{1}{N^4} [W^{(4)} - 4W^{(3)}W - 3(W^{(2)})^2 + 12W^{(2)}W^2 - 6W^4].$$

Now the  $\nu_{kl_1l_2\dots l_N}$ 's can be replaced by their values in terms of the  $\lambda_{kl_1l_2\dots l_N}$ 's, the details of which will be explained below, and it will be evident that any  $\nu_{kl_1l_2\dots l_N}$  is unaltered by a permutation of the  $l_i$ 's in its subscript. Taking account of this, the formulae (5) may be written in the expanded forms:

$$S_{11}(\bar{x}, m_n) = \frac{1}{N} [\nu_{1n} - \nu_{10}\nu_{0n}]$$

$$S_{12}(\bar{x}, m_n) = \frac{1}{N^2} [\nu_{2n} - \nu_{20}\nu_{0n} - 2\nu_{1n}\nu_{10} + 2\nu_{10}^2\nu_{0n}]$$

$$S_{12}(\bar{x}, m_n) = \frac{1}{N^2} [\nu_{1,2n} + (N-1)\nu_{1nn} - \nu_{10}\nu_{0,2n} - (N-1)\nu_{10}\nu_{0nn} - 2N\nu_{1n}\nu_{0n} + 2N\nu_{10}\nu_{0n}^2].$$

But, with no loss in generality, the origin may be taken at the population mean so that  $\lambda_1 = 0$ . In this case it will be found that  $\nu_{10} = 0$  and these formulae become:

$$S_{11}(\bar{x}, m_n) = \nu_{1n}/N$$

$$S_{21}(\bar{x}, m_n) = \frac{1}{N^2} [\nu_{2n} - \nu_{20}\nu_{0n}]$$

$$S_{12}(\bar{x}, m_n) = \frac{1}{N^2} [\nu_{1,2n} + (N-1)\nu_{1nn} - 2N\nu_{1n}\nu_{0n}]$$

$$S_{31}(\bar{x}, m_n) = \frac{1}{N^2} [\nu_{3n} - \nu_{30}\nu_{0n} - 3\nu_{1n}\nu_{20}]$$

$$(6) \quad S_{22}(\bar{x}, m_n) = \frac{1}{N^3} [\nu_{2,2n} + (N-1)\nu_{2nn} - 2N\nu_{2n}\nu_{0n} - \nu_{20}\nu_{0,2n} - (N-1)\nu_{20}\nu_{0nn} - 2N\nu_{1n}^2 + 2N\nu_{20}\nu_{0n}^2]$$

$$S_{13}(\bar{x}, m_n) = \frac{1}{N^3} [\nu_{1,3n} + 3(N-1)\nu_{1,2n,n} + (N-1)(N-2)\nu_{1nnn} - 3N\nu_{1,2n}\nu_{0n} - 3N(N-1)\nu_{1nn}\nu_{0n} - 3N\nu_{1n}\nu_{0,2n} - 3N(N-1)\nu_{1n}\nu_{0nn} + 6N^2\nu_{1n}\nu_{0n}^2].$$

These formulae are the second result used in the actual calculation of  $S_{ki}(\bar{x}, m_n)$ 's. One begins with them, putting in the particular value of  $n$  for the central moment in question. If for instance we wish to compute the product semi-invariants of the mean and variance in samples of  $N$ , we begin with the set of formulae:

$$S_{11}(\bar{x}, m_2) = \nu_{12}/N$$

$$(7) \quad S_{21}(\bar{x}, m_2) = \frac{1}{N^2} [\nu_{22} - \nu_{20}\nu_{02}]$$

$$S_{12}(\bar{x}, m_2) = \frac{1}{N^2} [\nu_{14} + (N-1)\nu_{122} - 2N\nu_{12}\nu_{02}],$$

etc.

The second step is to replace the product moments  $\nu_{k_1, l_2, \dots, l_N}$  which appear by their values in terms of the corresponding product semi-invariants. This process can perhaps be best explained by some examples.

Consider the complete calculation of  $S_{12}(\bar{x}, m_2)$ . From the expression for the fifth central moment in terms of semi-invariants:

$$\nu_5 = \lambda_5 + 10\lambda_3\lambda_2,$$

we can write the corresponding expression for product moments in terms of product semi-invariants

$$(8) \quad (\Sigma \nu_i \vartheta_i)^{(6)} \equiv (\Sigma \lambda_i \vartheta_i)^{(6)} + 10(\Sigma \lambda_i \vartheta_i)^{(3)}(\Sigma \lambda_i \vartheta_i)^{(2)}.$$

Then we get  $\nu_{14}$  by comparing coefficients of  $\frac{\vartheta_1 \vartheta_2^4}{1!4!}$  and  $\nu_{122}$  by comparing coefficients of  $\frac{\vartheta_1 \vartheta_2^2 \vartheta_3^2}{2!2!}$  in this identity. For an index as low as 5, these coefficients are readily picked out by inspection; for larger indices the use of Hammond operators reduces this to a mechanical routine.<sup>3</sup> In this case we have

$$D_3 D_2(14) = (12)(02) + (03)(11).$$

To the terms on the right the appropriate binomial coefficients must be applied giving

$$3(12)(02) + 2(03)(11).$$

The total of these coefficients is  $5 = \frac{5!}{4!1!}$ , a necessary check. Then multiplying these coefficients by  $10/5$ , we have

$$6\lambda_{12}\lambda_{02} + 4\lambda_{03}\lambda_{11}$$

for the required coefficients in the second term in (8). Thus

$$\nu_{14} = \lambda_{14} + (6\lambda_{12}\lambda_{02} + 4\lambda_{03}\lambda_{11}).$$

The two terms in parentheses arise from the same term in (8) and would both give rise to terms in  $\lambda_3\lambda_2$  in the final result if  $\lambda_{11}$  were not identically zero from (3). In practice all terms in which  $\lambda_{k1}$  is a factor are crossed out as they appear. Next

$$D_3 D_2(122) = 2(12)(02) + (111)(011) + 2(021)(11).$$

( $\lambda_{002} = \lambda_{02}$  ;  $\lambda_{012} = \lambda_{021}$ .) With the binomial, or multinomial coefficients attached, the right member is rewritten

$$6(12)(02) + 12(111)(011) + 12(021)(11).$$

<sup>3</sup> Cf. the author, loc. cit., p. 24.

The total of these coefficients is  $30 = \frac{5!}{2!2!1!1!}$ . Then multiplying each coefficient by  $10/30$ , we have

$$\nu_{122} = \lambda_{122} + (2\lambda_{12}\lambda_{02} + 4\lambda_{111}\lambda_{011} + 4\lambda_{012}\lambda_{11}).$$

Going on with the calculation of  $S_{12}(\bar{x}, m_2)$ :

$$\nu_{12} = \lambda_{12}, \quad \nu_{02} = \lambda_{02},$$

and then we have:

$$S_{12}(\bar{x}, m_2) = \frac{1}{N^2} [\{\lambda_{14} + (N-1)\lambda_{122}\} \\ + \{6\lambda_{12}\lambda_{02} + (N-1)(2\lambda_{12}\lambda_{02} + 4\lambda_{111}\lambda_{011}) - 2N\lambda_{12}\lambda_{02}\}].$$

The first set of terms within braces gives rise to terms in  $\lambda_5$ ; the second to terms in  $\lambda_3\lambda_2$ . Next

$$\begin{aligned} \lambda_{14} &= \frac{(N-1)(N^2-3N+3)}{N^3} \lambda_5 & \lambda_{111} &= -\frac{\lambda_3}{N} \\ \lambda_{122} &= \frac{2N-3}{N^3} \lambda_5 & \lambda_{02} &= \frac{N-1}{N} \lambda_2 \\ \lambda_{03} &= \frac{(N-1)(N-2)}{N^2} \lambda_3 & \lambda_{011} &= -\frac{\lambda_2}{N} \\ \lambda_{021} &= -\frac{(N-2)}{N^2} \lambda_3 \end{aligned}$$

This table of values will be of frequent use in further calculations of  $S_{kl}$ 's. Giving the values of both  $\lambda_{111}$  and  $\lambda_{011}$  here, was unnecessary duplication.

Now only the final reduction is to be carried out. We obtain

$$S_{12}(\bar{x}, m_2) = \frac{N-1}{N^4} [(N-1)\lambda_5 + 4N\lambda_3\lambda_2].$$

This result of order 3 and of weight 5 follows a quite mechanical procedure and is quite brief. The length of the algebraic computations required grows rapidly as the weight is increased but for weights no greater than 10 undue labor is not required. For greater weights only time and patience is required to get results if they are needed. It is to be noted that by this method one may calculate individual terms in the result without doing any of the work required for the remaining terms and that one may readily shorten the work by getting results to a desired degree of approximation with respect to powers of  $1/N$ .

There follows a table of the results so far calculated.

For  $n = 2$ :

$$S_{11} = \frac{N-1}{N} \lambda_3$$

$$S_{21} = \frac{N-1}{N^2} \lambda_4$$

$$S_{12} = \frac{N-1}{N^4} [(N-1)\lambda_5 + 4N\lambda_3\lambda_2]$$

$$S_{31} = \frac{N-1}{N^4} \lambda_5$$

$$S_{22} = \frac{N-1}{N^5} [(N-1)\lambda_6 + 4N(\lambda_4\lambda_2 + \lambda_3^2)]$$

$$S_{13} = \frac{N-1}{N^6} [(N-1)^2\lambda_7 + 12N(N-1)\lambda_5\lambda_2 + 4N(5N-7)\lambda_4\lambda_3 + 24N^2\lambda_3\lambda_2^2].$$

It is not difficult to see that in general

$$S_{k1}(\bar{x}, m_2) = \frac{N-1}{N^{k+1}} \lambda_{k+2}.$$

For  $n = 3$ :

$$S_{11} = \frac{(N-1)(N-2)}{N^3} \lambda_4$$

$$S_{21} = \frac{(N-1)(N-2)}{N^4} \lambda_5$$

$$S_{12} = \frac{(N-1)(N-2)}{N^6} [(N-1)(N-2)\lambda_7 + 9N(N-2)\lambda_5\lambda_2 \\ + 27N(N-2)\lambda_4\lambda_3 + 18N^2\lambda_3\lambda_2^2]$$

$$S_{31} = \frac{(N-1)(N-2)}{N^6} \lambda_6$$

$$S_{22} = \frac{(N-1)(N-2)}{N^7} [(N-1)(N-2)\lambda_8 + 9N(N-2)\lambda_6\lambda_2 \\ + 36N(N-2)\lambda_5\lambda_3 + 27N(N-2)\lambda_4^2 + 18N^2\lambda_4\lambda_2^2 + 36N^2\lambda_3^2\lambda_2]$$

$$S_{13} = \frac{(N-1)(N-2)}{N^{10}} [N(N-1)^2(N-2)^2\lambda_{10} \\ + 9(N-1)(3N^4 - 12N^3 + 12N^2 - 5N + 5)\lambda_8\lambda_2 \\ + 27N(4N^4 - 21N^3 + 36N^2 - 20N + 3)\lambda_7\lambda_3 \\ + 27N^2(N-2)^2(7N-11)\lambda_6\lambda_4 + 54N^3(N-2)(4N-7)\lambda_6\lambda_3^2]$$

$$\begin{aligned}
& + 27N^2(N-2)^2(4N-7)\lambda_1^2 + 54N^3(N-2)(23N-50)\lambda_4\lambda_5\lambda_6 \\
& + 162N^3(N-2)(5N-12)\lambda_1^2\lambda_2 + 54N^2(29N^2-126N+140)\lambda_4\lambda_5^2 \\
& + 108N^4(5N-12)\lambda_4\lambda_5^2 + 324N^4(5N-12)\lambda_5^2\lambda_6^2].
\end{aligned}$$

For  $n = 4$ :

$$S_{11} = \frac{N-1}{N^4} [(N^2-3N+3)\lambda_5 + 6N(N-1)\lambda_2\lambda_3]$$

$$S_{21} = \frac{N-1}{N^5} [(N^2-3N+3)\lambda_6 + 6N(N-1)(\lambda_4\lambda_5 + \lambda_5^2)]$$

$$\begin{aligned}
S_{12} = \frac{N-1}{N^8} [(N-1)(N^2-3N+3)^2\lambda_9 \\
& + 4N(N^2-3N+3)(7N^2-18N+15)\lambda_7\lambda_8 \\
& + 4N(N^2-3N+3)(19N^2-66N+63)\lambda_8\lambda_9 \\
& + 4N(29N^4-195N^3+537N^2-639N+351)\lambda_5\lambda_4 \\
& + 12N^2(17N^3-71N^2+117N-69)\lambda_5\lambda_5^2 \\
& + 24N^2(35N^3-173N^2+309N-189)\lambda_4\lambda_5\lambda_2 \\
& + 72N^2(N-2)^2(3N-5)\lambda_5^3 + 96N^3(4N^2-9N+6)\lambda_5\lambda_5^2]
\end{aligned}$$

$$S_{11} = \frac{N-1}{N^6} [(N^2-3N+3)\lambda_7 + 6N(N-1)\lambda_5\lambda_2 + 18N(N-1)\lambda_4\lambda_3]$$

$$\begin{aligned}
S_{22} = \frac{N-1}{N^9} [(N-1)(N^2-3N+3)^2\lambda_{10} \\
& + 4N(N^2-3N+3)(7N^2-18N+15)\lambda_8\lambda_2 \\
& + 8N(N^2-3N+3)(13N^2-42N+39)\lambda_7\lambda_3 \\
& + 12N(16N^4-106N^3+285N^2-360N+180)\lambda_5\lambda_4 \\
& + 12N^2(17N^3-71N^2+117N-69)\lambda_5\lambda_5^2 \\
& + 4N(29N^4-195N^3+537N^2-693N+351)\lambda_5^3 \\
& + 48N^2(26N^3-125N^2+213N-129)\lambda_5\lambda_3\lambda_2 \\
& + 24N^2(35N^3-173N^2+309N-189)\lambda_4^2\lambda_2 \\
& + 24N^2(62N^3-326N^2+597N-369)\lambda_4\lambda_5^2 \\
& + 96N^3(4N^2-9N+6)\lambda_4\lambda_5^2 + 288N^3(4N^2-9N+6)\lambda_5^2\lambda_5^2].
\end{aligned}$$



# ON THE NON-EXISTENCE OF TESTS OF "STUDENT'S" HYPOTHESIS HAVING POWER FUNCTIONS INDEPENDENT OF $\sigma$

BY GEORGE B. DANTZIG

**1. Introduction.** Consider a system of  $n$  random variables  $x_1, x_2, \dots, x_n$  where each is known to be normally distributed about the same but unknown mean,  $\xi$ , and with the same, but also unknown standard deviation  $\sigma$ . The assumption,  $H_0$ , that  $\xi$  has some specified value,  $\xi_0$ , e.g.  $\xi_0 = 0$ , while nothing is assumed about  $\sigma$ , is known as the "Student" Hypothesis. Two aspects of the hypothesis  $H_0$  have been already studied extensively. If the alternatives with respect to which it is desired to test  $H_0$  assume specifically that  $\xi > \xi_0$ , (or  $\xi < 0$ ), then we have the so-called asymmetric case of "Student's Hypothesis" and it is known, [1], that there exists a uniformly most powerful test of  $H_0$ . This consists in the rule, originally suggested by "Student," of rejecting  $H_0$  whenever

$$(1) \quad t = \frac{\bar{x} - \xi_0}{S} \sqrt{n-1} > t_\alpha,$$

where  $\bar{x}$  and  $S$  denote the mean and the standard deviation of the observed  $x_i$ 's and  $t_\alpha$  is taken, for example, from Fisher's Tables [2] with his  $P = 2\alpha$ . In other words  $t_\alpha$  is such that

$$(2) \quad P\{t > t_\alpha \mid H_0\} = \alpha,$$

where  $\alpha$  is the chosen level of significance. In accordance with the definition of the uniformly most powerful test, whenever any other rule,  $R$ , offered to test the same hypothesis  $H_0$  has the same probability  $\alpha$  of  $H_0$  being rejected when it is true, the power of this alternative test cannot exceed that of "Student's" Test. In other words, if it happens that the true value of  $\xi$  is not equal to  $\xi_0$  but is greater, then the probability of this circumstance being detected by "Student's" test is at least equal to that corresponding to the rule  $R$ .

If the set of alternative hypotheses is not limited to those specifying the value of  $\xi$  either greater or smaller than  $\xi_0$ , but includes both those categories, then it is known, [1], that there is no uniformly most powerful test of the hypothesis,  $H_0$ . However in this case there exists a slightly different test, also based on "Student's" criterion  $t$ , possessing the remarkable property of being unbiased of type  $B_1$ , [3]. The test, in common use for a long time, consists in rejecting  $H_0$  when

$$(3) \quad |t| > t_\alpha,$$

with  $t_\alpha$  being taken again from Fisher's tables, this time corresponding to his  $P = \alpha$ , where  $\alpha$  is the chosen level of significance.

In order to describe the optimum property of this test we must use the concept of the power function of a test, [3]. Denote by  $\beta(\xi, \sigma)$  the probability of the hypothesis  $H_0$  being rejected when  $\xi$  and  $\sigma$  are the true mean and the true standard error of the observable  $x_i$ 's. The function  $\beta(\xi, \sigma)$  is just what is called the power function of the test. If we substitute  $\xi = \xi_0$ , then we shall have  $\beta(\xi_0, \sigma) = \alpha$  irrespective of the value of  $\sigma$ . Now the optimum property of "Student's" test mentioned above consists in that (1) its power function has a minimum at  $\xi = \xi_0$  and this is true whatever be the value of  $\sigma$ , (2) whatever be any other test of the same hypothesis which has the same level of significance  $\alpha$  and has property (1), its power function  $\beta'(\xi, \sigma)$  cannot exceed that of "Student's" test.

These two properties, demonstrating the excellence of the criterion suggested by "Student," fully justify the general confidence in the test as described above, or in its extended form where it is applied to two or more samples. However, it is known that "Student's" test in both its forms,  $t > t_\alpha$ , and  $|t| > t_\alpha$ , has one very undesirable property which causes great difficulties in various problems of rational planning of experiments.

One of the most important questions to have in mind when planning an experiment is: What is the probability that the experiment and the subsequent statistical test will detect a difference or effect when it actually exists? If we perform an experiment and then apply some statistical analysis to test "Student's" hypothesis that  $\xi = \xi_0$ , we do hope that, if the actual value of  $\xi$  is different from  $\xi_0$ , the test will discover this circumstance. But apart from mere hope, it is desirable to take precautions so that when the *difference*,  $\xi - \xi_0 = \Delta$ , has some appreciable value, the chance of the hypothesis  $H_0$  being rejected will be reasonably large. This may be done by calculating the value of the power function  $\beta(\xi, \sigma)$  corresponding to the value  $\xi = \xi_0 + \Delta$ . And here we come to the unfortunate property of "Student's" test.

Although the form of the power function of "Student's" test is known and tabled [4], [5], [6], [7], there are occasionally considerable difficulties in applying these tables, because it appears that the values  $n$  and  $\Delta$  are not all its arguments, for it also depends on  $\sigma$ . Consequently in order to have an idea of the probability that the test will detect the falsehood of the hypothesis  $H_0$  that  $\xi = \xi_0$  when actually  $\xi = \xi_0 + \Delta$  we need not only the knowledge of  $n$  but also a likely value of  $\sigma$ . The latter is known accurately only in exceptional cases and then in those cases one would apply a test which is different from "Student's" test. Usually we have only a vague notion of the magnitude of  $\sigma$  and accordingly the tables of  $\beta(\xi, \sigma)$  may be used to obtain a rough idea as to whether the arrangement of the experiment planned is satisfactory or not. Frequently we have no idea of what may be the values of  $\sigma$ .

To Dr. P. L. Hsu is due the idea of looking for tests, the power of which is independent of the parameters unspecified by the hypothesis tested. In an

unpublished paper, he proved among other things that the  $\lambda$  test of the general linear hypothesis is the most powerful of all those, the power function of which depends on the same argument as that of the  $\lambda$  test and not on other parameters. The above circumstances suggest the following problem: to see whether it is possible to devise a test of "Student's" hypothesis such that its power function would be independent of  $\sigma$ . If such a test could be devised and proved to be reasonably powerful then the tables of its power function could be used for the purpose of planning experiments.

The purpose of the present paper is to show that no such test exists and, consequently, this negative result implies in still another way that it is impossible to improve on the test originally suggested by "Student."

**2. Statement of the Problem.** The problem of finding a test whose power function is independent of  $\sigma$  is equivalent to finding a critical region  $w$  such that the value of the power function

$$(4) \quad \beta(\xi, \sigma) = P\{E \in w \mid \xi, \sigma\}$$

for any fixed  $\xi$  is independent of the value of  $\sigma$ , where  $E$  denotes the sample point  $(x_1, x_2, \dots, x_n)$ . We shall show specifically that if this is the case, then the power function is also independent of  $\xi$ ; so that the test will reject the hypothesis tested with the same frequency independently of whether it be correct or wrong.

**3. THEOREM.** *If there exists a region  $w$  such that, whatever be the value of  $\sigma$ ,*

$$(5) \quad \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \int \dots \int_w e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \xi_0)^2} dx_1 dx_2 \dots dx_n = \alpha$$

$$(6) \quad \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \int \dots \int_w e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \xi_1)^2} dx_1 dx_2 \dots dx_n = \beta,$$

where  $\xi_0 \neq \xi_1$ ,  $\alpha, \beta$  are constants, then

$$(7) \quad \alpha = \beta.$$

A region  $w$  is called *similar* [1] to the whole sample space,  $W$ , of size  $\alpha$ , with respect to a set of elementary probability laws  $p(E \mid \theta)$  given in terms of a parameter  $\theta$ , if  $P\{E \in w \mid \theta\} = \alpha$ , whatever be the value of  $\theta$ . Essentially, then, the region,  $w$ , above is a similar region with respect to two different sets of elementary laws each being given parametrically in terms of the parameter  $\sigma$ .

Denote by  $w_r$  the portion of the surface of the hypersphere,  $\sum_{i=1}^n (x_i - \xi_0)^2 = r^2$ , which is common to  $w$ , and let the total surface be denoted by  $W_r$ . Neyman and Pearson have shown [1], that a necessary and sufficient condition that  $w$  be a similar region, in the above case, is that, whatever be  $r$ , the probability

that the sample point  $E$  will fall on the subsurface  $w_r$ , when it is known that the sample point lies on the surface  $W_r$  is  $\alpha$ , i.e.

$$(8) \quad P\{E \in w_r \mid (E \in W_r)(\xi = \xi_0)\} = \alpha$$

for all  $r$ .

In a similar manner let  $w_\rho$  denote the portion of the surface of the hypersphere  $\sum_{i=1}^n (x_i - \xi_i)^2 = \rho^2$  common to  $w$ , and let the total surface be denoted by  $W_\rho$ . Since  $w$  is similar to the set of probability laws indicated in (6), we have also

$$(9) \quad P\{E \in w_\rho \mid (E \in W_\rho)(\xi = \xi_1)\} = \beta$$

for all  $\rho$ .

Since on the surface  $W_r$ , the elementary probability law,

$$(10) \quad \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \xi_0)^2} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{r^2}{2\sigma^2}},$$

is constant, we see that an equivalent statement of (8) is that *the hyper-area of  $w_r$  is a constant proportion,  $\alpha$ , of the total hyper-area  $W_r$* . Similarly, from (9), we have that *the hyper-area of  $w_\rho$  is a constant proportion,  $\beta$ , of the area of the hypersurface  $W_\rho$ , whatever be the values of  $r$  and  $\rho$* .

Consider the transformation which expresses  $x_1, x_2, \dots, x_n$  in terms of generalized polar coordinates with pole at the point  $(\xi_0, \xi_0, \dots, \xi_0)$ , i.e.

$$(11) \quad \begin{aligned} x_1 - \xi_0 &= r \cos \theta_2 \cos \theta_3 \dots \cos \theta_{n-2} \cos \theta_{n-1} \cos \theta_n \\ x_2 - \xi_0 &= r \cos \theta_2 \cos \theta_3 \dots \cos \theta_{n-2} \cos \theta_{n-1} \sin \theta_n \\ x_3 - \xi_0 &= r \cos \theta_2 \cos \theta_3 \dots \cos \theta_{n-2} \sin \theta_{n-1} \end{aligned}$$

$$x_{n-1} - \xi_0 = r \cos \theta_2 \sin \theta_3$$

$$x_n - \xi_0 = r \sin \theta_2$$

Let  $\Delta$  be the Jacobian of the transformation:

$$(12) \quad |\Delta| = r^{n-1} \prod_{i=2}^n \cos^i \theta_{n+2-i} = r^{n-1} T(\theta_i).$$

Consider also a transformation which expresses  $(x_1, x_2, \dots, x_n)$  in terms of polar coordinates, the point  $(\xi_1, \xi_1, \dots, \xi_1)$  being pole. It may be obtained by replacing in (11),  $\xi_0$  by  $\xi_1$ ,  $r$  by  $\rho$ , and  $\theta_i$  by  $\bar{\theta}_i$ . The Jacobian of this transformation is given by  $|\bar{\Delta}| = \rho^{n-1} T(\bar{\theta}_i)$ .

We are now able to express the *hyper-area* of  $W_r$ :

$$(13) \quad \iint |\Delta| d\theta_2 d\theta_3 \dots d\theta_n = r^{n-1} \iint_{w_r} T(\theta_i) d\theta_2 d\theta_3 \dots d\theta_n = Kr^{n-1},$$

where the integral  $K > 0$  is a constant independent of  $r$ . Similarly the hyper-area of  $W_\rho$  is  $K\rho^{n-1}$ , where  $K$  is the same as in (13). According to (8) and (9) we have, now

$$(14) \quad \int \int_{w_r} |\Delta| d\theta_2 d\theta_3 \dots d\theta_n = \alpha \cdot K \cdot r^{n-1},$$

$$(15) \quad \int \int_{w_\rho} |\bar{\Delta}| d\bar{\theta}_2 d\bar{\theta}_3 \dots d\bar{\theta}_n = \beta \cdot K \cdot \rho^{n-1}.$$

Let us consider the distances between the three points:  $(x_1, x_2, \dots, x_n)$ ,  $(\xi_0, \xi_0, \dots, \xi_0)$ , and  $(\xi_1, \xi_1, \dots, \xi_1)$ . The distances of the first point to the second point and to the third point we have already denoted by  $r$  and  $\rho$ . Let the distance between last two be  $L$ , then, since the sum of two sides is at least equal to the third side of a triangle, we have

$$(16) \quad r \leq \rho + L, \quad \rho \leq r + L, \quad \text{where } L = \sqrt{N} |\xi_0 - \xi_1|.$$

Let  $\varphi(t) \geq 0$  be an arbitrary monotonic nonincreasing function of  $t$ , such that the product  $t^{n-1}\varphi(t)$  is integrable from 0 to  $+\infty$ . Since  $\varphi(t)$  is a decreasing function it follows from (16) that

$$(17) \quad \varphi(r) \geq \varphi(\rho + L) \quad \text{and} \quad \varphi(\rho) \geq \varphi(r + L).$$

Consider the integral  $I$ :

$$(18) \quad I = \int \int \varphi(r) dx_1 dx_2 \dots dx_n.$$

We shall express it in terms of the variables  $r, \theta_2, \dots, \theta_n$  and also in terms of  $\rho, \bar{\theta}_2, \dots, \bar{\theta}_n$  and compare the results. Thus

$$\begin{aligned} I &= \int \int_w |\Delta| \varphi(r) dr d\theta_2 \dots d\theta_n \\ (19) \quad &= \int_0^\infty \varphi(r) dr \int \int_{w_r} |\Delta| d\theta_2 \dots d\theta_n \\ &= \alpha \cdot K \cdot \int_0^\infty r^{n-1} \varphi(r) dr. \end{aligned}$$

Also we have by (16)

$$I = \int \int |\bar{\Delta}| \varphi(\rho) d\rho d\bar{\theta}_2 \dots d\bar{\theta}_n$$

$$\begin{aligned}
 (20) \quad & \geq \int \int_{\mathbf{w}} |\bar{\Delta}| \varphi(\rho + L) d\rho d\bar{\theta}_2 \dots d\bar{\theta}_n \\
 & \geq \int_0^\infty \varphi(\rho + L) d\rho \int \int_{\mathbf{w}} |\bar{\Delta}| d\bar{\theta}_2 \dots d\bar{\theta}_n
 \end{aligned}$$

and consequently

$$(21) \quad I \geq \beta \cdot K \int_0^\infty \rho^{n-1} \varphi(\rho + L) d\rho.$$

Since  $K > 0$ , we have from (19) and (21)

$$(22) \quad \alpha/\beta \geq \int_0^\infty t^{n-1} \varphi(t + L) dt / \int_0^\infty t^{n-1} \varphi(t) dt.$$

By interchanging  $\rho$  and  $r$  in (18), (19), (20), and (21) we have also

$$(23) \quad \beta/\alpha \geq \int_0^\infty t^{n-1} \varphi(t + L) dt / \int_0^\infty t^{n-1} \varphi(t) dt.$$

Let us set in (22) and (23),  $\varphi(t) = e^{-pt}$  and  $\varphi(t + L) = e^{-pL} e^{-pt}$  where  $p > 0$  is arbitrary. Then

$$(24) \quad \alpha/\beta \geq e^{-pL} \quad \text{and} \quad \beta/\alpha \geq e^{-pL}.$$

Since (24) holds for all  $p > 0$ , let  $p$  approach zero. Then  $\lim e^{-pL} = 1$ , and the above inequalities can hold only if

$$(25) \quad \alpha = \beta, \quad \text{Q.E.D.}$$

It is of interest to note that there do exist regions such that the power function is independent of both  $\xi$  and  $\sigma$ . For example, let  $S_n$  be the standard deviation of the observed values  $(x_1, x_2, \dots, x_n)$  and let  $S_{n-1}$  be the standard deviation of the values  $(x_1, x_2, \dots, x_{n-1})$ , then the region  $w$  given by all points  $(x_1, x_2, \dots, x_n)$  which satisfy the inequality  $(S_{n-1}/S_n) \geq C$  is such a region, i.e.

$$(26) \quad P\{(S_{n-1}/S_n) \geq C \mid \xi, \sigma\}$$

is constant, whatever be the values of  $\xi$  and  $\sigma$ . Such regions are, however, unsuitable for testing "Student's" hypothesis  $\xi = \xi_0$ , because they will reject this hypothesis when it is wrong and when it is correct with equal frequency.

The author is indebted to Professor J. Neyman for assistance in preparing the present paper.

#### REFERENCES

- [1] J. NEYMAN AND E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc. London*, Vol. 231(1933), pp. 289-337.
- [2] R. A. FISHER, *Statistical methods for research workers*. Oliver & Boyd, 7th edition, London, 1938.

- [3] J. NEYMAN, "Sur la vérification des hypothèses statistiques composées," *Bull. Soc. Math. de France*, T. 63 (1935), pp. 246-266.
- [4] S. KOŁODZIEJCZYK, "Sur l'erreur de la deuxième catégorie dans le problème de 'Student,'" *Compte Rendus*, T. 197(1933), p. 814.
- [5] J. NEYMAN with co-operation of K. IWASZKIEWICZ AND S. KOŁODZIEJCZYK, "Statistical problems in agricultural experimentation," *Suppl. Jour. Roy. Stat. Soc.*, Vol. II(1935), pp. 107-80.
- [6] J. NEYMAN AND B. TOKARSKA, "Errors of the second kind in testing 'Student's' hypothesis," *Jour. Am. Stat. Ass.*, Vol. 31(1936), pp. 318-26.
- [7] P. C. TANG, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res. Memoirs*, Vol. II(1938), pp. 126-57.

UNIVERSITY OF CALIFORNIA,  
BERKELEY, CALIFORNIA.

# A METHOD FOR RECURRENT COMPUTATION OF ALL THE PRINCIPAL MINORS OF A DETERMINANT, AND ITS APPLICATION IN CONFLUENCE ANALYSIS

BY OLAV REIERSØL

1. Recurrent computation of all the principal minors of a determinant. The formulae which I develop in this paper have been worked out for use in statistical confluence analysis. By means of recurrent computation they shorten considerably the amount of work required to compute all principal minors of a square matrix. Originally I elaborated this method as a simplification of one given by Frisch (not published).

Subsequently I found that the method could more easily be deduced from the pivotal method. This method has been described, for example, by Whittaker and Robinson [5] and by Aitken [1].

Let us consider a square  $n$ -rowed matrix

$$(1) \quad \begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{array}$$

Let the adjoint of this matrix be  $\|p_{ij}\|$  and let us denote its determinant value by  $D_{12 \dots n}$ .

Then we have the following identity

$$(2) \quad \begin{vmatrix} p_{n-1, n-1} & p_{n-1, n} \\ p_{n, n-1} & p_{n, n} \end{vmatrix} = D_{12 \dots n} D_{12 \dots n-2}.$$

As Aitken points out, the pivotal method is based upon this identity.

Next consider the following matrix which is formed from the matrix (1) by striking out the  $n$ th row and the  $(n - 1)$ th column:

$$(3) \quad \begin{array}{cccc} a_{11} & \cdots & a_{1, n-2} & a_{1, n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n-2, 1} & \cdots & a_{n-2, n-2} & a_{n-2, n} \\ a_{n-1, 1} & \cdots & a_{n-1, n-2} & a_{n-1, n} \end{array}$$



Let us denote its adjoint by  $|| q_{ij} ||$ , its determinant value by  $A_{12\dots n} = -p_{n-1,n}$ . The determinant

$$\begin{vmatrix} a_{11} & \dots & a_{1,n-2} & a_{1,n-1} \\ a_{n-2,1} & \dots & a_{n-2,n-2} & a_{n-2,n-1} \\ a_{n,1} & \dots & a_{n,n-2} & a_{n,n-1} \end{vmatrix} = -p_{n,n-1},$$

we shall denote by  $B_{12\dots n}$ .

The identity (2) can now be written

$$(2') \quad D_{12\dots n} = \frac{D_{12\dots n-2,n} D_{12\dots n-2,n-1} - A_{12\dots n} B_{12\dots n}}{D_{12\dots n-2}}.$$

If we apply the identity (2) to the matrix (3) we get

$$\begin{vmatrix} q_{n-2,n-2} & q_{n-2,n-1} \\ q_{n-1,n-2} & q_{n-1,n-1} \end{vmatrix} = A_{12\dots n} D_{12\dots n-3},$$

which may also be written

$$(4) \quad A_{12\dots n} = \frac{A_{12\dots n-3,n-1,n} D_{12\dots n-2} - A_{12\dots n-3,n-2,n} B_{12\dots n-1}}{D_{12\dots n-3}}.$$

To simplify the notation we will not write the affixes present, but write the affixes not present in inverted parentheses. Then our formulae (2') and (4) can be written

$$D = \frac{D_{)n-1}(D_{)n(} - AB}{D_{)n-1,n(}},$$

$$A = \frac{A_{)n-2}(D_{)n-1,n(} - A_{)n-1}(B_{)n(}}{D_{)n-2,n-1,n(}}.$$

In an analogous way we get

$$B = \frac{B_{)n-2}(D_{)n-1,n(} - B_{)n-1}(A_{)n(}}{D_{)n-2,n-1,n(}}.$$

We may apply these formulae to an arbitrary principal minor  $D_{v_1 v_2 \dots v_k}$ . Let us now denote  $D_{v_1 v_2 \dots v_k}$  by  $D$  and denote the absence of one or more of the numbers  $v_1, v_2, \dots, v_k$  by placing them into inverted parentheses. We then have the formulae:

$$(5a) \quad A = \frac{A_{)v_{k-2}(D_{)v_{k-1},v_k(} - A_{)v_{k-1}(B_{)v_k(}}{D_{)v_{k-2},v_{k-1},v_k(}},$$

$$(5b) \quad B = \frac{B_{)v_{k-2}(D_{)v_{k-1},v_k(} - B_{)v_{k-1}(A_{)v_k(}}{D_{)v_{k-2},v_{k-1},v_k(}},$$

$$(5c) \quad D = \frac{D_{)v_{k-1}(D_{)v_k(} - AB}{D_{)v_{k-1},v_k(}}.$$

By means of these formulae we can recurrently compute all principal minors. We begin with  $D_i = a_{ii}$ ,  $i = 1, 2 \dots n$ ,  $A_{ij} = a_{ij}$ ,  $B_{ij} = a_{ji}$ , where  $i < j$ . Then we compute the  $D$ 's with two affixes,

$$D_{ij} = D_i D_j - A_{ij} B_{ij},$$

and then the quantities  $A$ ,  $B$ ,  $D$  with three affixes,

$$A_{ijk} = A_{jk} D_i - A_{ik} B_{ij}$$

$$B_{ijk} = B_{jk} D_i - B_{ik} A_{ij}$$

$$D_{ijk} = \frac{D_{ik} D_{ij} - A_{ijk} B_{ijk}}{D_i}, \quad i < j < k.$$

Then we compute the quantities  $A$ ,  $B$ ,  $D$  with four affixes, and so on.

If we carry through the computations without dropping any figures we have as a control that all divisions will be exact without remainder. If we are dropping figures we can control the result by computing the determinant  $D_{12\dots n}$  in another way. If we wish to control the computation before it is completed, we may use our recurrence formulae on the matrix which we get from the original matrix when the rows and the columns are subjected to the same permutation. For example we can reverse the order of the rows and columns. Then we can control the  $(k - 1)$  rowed minors before computing the  $k$ -rowed minors.

If all the  $D$ 's are different from zero, we may reduce the necessary number of multiplications and divisions in the following way. We introduce the following notations:

$$\begin{aligned} d &= \frac{D}{D_{)v_k(}} \\ a &= \frac{A}{D_{)v_{k-1}, v_k(}} & b &= \frac{B}{D_{)v_{k-1}, v_k(}} \\ c &= -\frac{b}{d_{)v_k(}} \end{aligned}$$

Substituting in (5), we get the following system of recurrence formulae:

$$(6a) \quad a = a_{)v_{k-2}(} + a_{)v_{k-1}(} c_{)v_k(}$$

$$(6b) \quad b = b_{)v_{k-2}(} + a_{)v_k(} c_{)v_{k-1}(}$$

$$(6c) \quad c = -\frac{b}{d_{)v_k(}}$$

$$(6d) \quad d = d_{)v_{k-1}(} + ac$$

$$(6e) \quad D = D_{)v_k(} d.$$

An affix  $v_k$  on a letter indicates the deletion of the last row and column in the determinants making up the definition of that letter, even though those determinants are of lower order than  $v_k$ . Similarly, an affix  $v_{k-1}$  indicates the deletion of the next to the last row and column.

The  $a$ 's with two affixes in these formulae are identical with the elements  $a_{ij}$  of the matrix (1) where  $i < j$ . Further,  $b_{ij} = a_{ji}$ ,  $i < j$ ,  $d_i = a_{ii}$ . Applying the recurrence formulae (6) we start with these values.

If the matrix (1) is symmetric, i.e. if  $a_{ij} = a_{ji}$ , then we get

$$B_{v_1 v_2 \dots v_k} = A_{v_1 v_2 \dots v_k}$$

and

$$b_{v_1 v_2 \dots v_k} = a_{v_1 v_2 \dots v_k}.$$

In this case we can therefore replace  $B$  by  $A$  in the formulae (5) and replace  $b$  by  $a$  in the formulae (6).

*Numerical example.* Let us compute all the scatterances in the constructed example given by Frisch, [3, p. 121]. The correlation matrix in this example is:

1.000000	-0.121551	0.656809	0.752502	-0.224549
-0.121551	1.000000	0.657698	-0.732862	0.212165
0.656809	0.657698	1.000000	0.014385	-0.040183
0.752502	-0.732862	0.014385	1.000000	-0.280223
-0.224549	0.212165	-0.040183	-0.280223	1.000000

Using our recurrence formulae (6) we get the following table:

	$a$	$c$	$d$	$D$
12	-0.121 551	0.121 551	0.985 225	0.985 225
13	0.656 809	-0.656 809	0.568 602	0.568 602
23	0.657 698	-0.657 698	0.567 433	0.567 433
14	0.752 502	-0.752 502	0.433 741	0.433 741
24	-0.732 862	0.732 862	0.462 913	0.462 913
34	0.014 385	-0.014 385	0.999 793	0.999 793
15	-0.224 549	0.224 549	0.949 578	0.949 578
25	0.212 165	-0.212 165	0.954 986	0.954 986
35	-0.040 183	0.040 183	0.998 385	0.998 385
45	-0.280 223	0.280 223	0.921 475	0.921 475
123	0.737 534	-0.748 594	0.016 489	0.016 245
124	-0.641 395	0.651 014	0.016 184	0.015 945
134	-0.479 865	0.843 938	0.028 765	0.016 356
234	0.496 387	-0.874 794	0.028 677	0.016 272
125	0.184 871	-0.187 643	0.914 888	0.901 371
135	0.107 303	-0.188 714	0.929 328	0.528 418
235	-0.179 723	0.316 730	0.898 062	0.509 590
145	-0.111 249	0.256 487	0.921 044	0.399 495
245	-0.124 785	0.269 457	0.921 272	0.426 516
345	-0.279 645	0.279 703	0.920 167	0.919 977

	<i>a</i>	<i>c</i>	<i>d</i>	<i>D</i>
1234	0.000 279	-0.016 6	0.016 179	0.000 262 83
1235	-0.031 090	1.885 5	0.856 268	0.013 910
1245	0.009 105	-0.562 6	0.909 766	0.014 506
1345	-0.020 692	0.719 35	0.914 443	0.014 957
2345	0.032 486	-1.132 8	0.861 262	0.014 014
12345	0.009 621	-0.594 7	0.850 546	0.000 223 55

**2. Computation of the coefficients of the characteristic polynomial of a matrix.** The characteristic polynomial of the matrix (1) is

$$P(\lambda) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{vmatrix}$$

$$= P_n - P_{n-1}\lambda + P_{n-2}\lambda^2 - \dots + (-1)^n \lambda^n.$$

As is well known, the coefficient  $P_k$  can be calculated as the sum of all the  $k$ -rowed principal minors of the matrix (1). Our method of computing all the principal minors of a matrix therefore gives us as a by-product a method of computing the coefficients of the characteristic polynomial. Another method for the determination of these coefficients has been given by Paul Horst [4].

We may obtain a comparison between the work of computation entailed by the two methods by calculating the number of multiplications and divisions necessary when using one or the other method. If our recurrence formulae (6) are used, two multiplications and one division are necessary for computing a 2-rowed minor, and 4 multiplications and one division for every minor with 3 or more rows. Consequently the total number of multiplications and divisions will be:

$$S_n = 3 \binom{n}{2} + 5 \sum_{k=3}^n \binom{n}{k}$$

$$= 5 \cdot 2^n - (n^2 + 4n + 5).$$

On using Horst's method, the number of necessary multiplications and divisions will be found to be

$$H_n = (\tfrac{1}{2}n - 1)n^3 + \tfrac{1}{2}n^3 + \tfrac{1}{2}(n - 1)(n + 2)$$

$$H_n = \tfrac{1}{2}(n - 1)(n^3 + n + 2) \quad n \text{ even,}$$

$$H_n = \tfrac{1}{2}(n - 1)(n^3 + n^2 + n + 2) \quad n \text{ odd.}$$

When  $n = 2, 3, \dots, 12$ ,  $S_n$  and  $H_n$  acquire the following values:

$n$	$S_n$	$H_n$
2	3	6
3	14	41
4	43	105
5	110	314
6	255	560
7	558	1203
8	1179	1827
9	2438	3284
10	4975	4554
11	10070	7325
12	20283	9581

We see that our method of computing the coefficients of the characteristic polynomial involves less calculation when  $n < 10$ , while Horst's method is superior when  $n \geq 10$ .

If our purpose is to find the characteristic roots of the matrix we can do this with less amount of computation without first finding the coefficients of the characteristic polynomial. See Aitken, [2].

**3. Applications in confluence analysis.** The confluence analysis of Frisch is set forth in his book: "Statistical Confluence Analysis by Means of Complete Regression Systems," [3].

The main method of this book is the "bunch analysis," which includes the computation of the adjoints of the correlation matrices of all sets of variates contained in the total set. In section 1, Frisch has described a preliminary analysis by means of scatterances. The scatterances are the principal minors of the correlation matrix of the total set of variates. If we carry through such an analysis, the recurrence formulae of section 1 of this paper will give a rapid method for the calculation of all the scatterances.

Another application of the computation of all the scatterances arises in the determination of the correct time lags between variates in a structural equation. This problem will be treated in a paper on confluence analysis which will appear in the near future.

#### REFERENCES

- [1] A. C. AITKEN, "On the Evaluation of Determinants, the Formation of their Adjugates, and the Practical Solution of Simultaneous Linear Equations," *Proc. of Edinburgh Math. Soc.*, Second Series. Vol. 3 (1932-33), p. 207.
- [2] A. C. AITKEN, "Studies in Practical Mathematics II," *Proc. Royal Soc. of Edinburgh*, Vol. 57 (1936-37), p. 269.
- [3] RAGNAR FRISCH, "Statistical Confluence Analysis by Means of Complete Regression Systems," *Publication no. 5 from the University Institute of Economics*, Oslo, 1934.
- [4] PAUL HORST, "A Method for Determining the Coefficients of a Characteristic Equation," *Annals of Math. Stat.*, Vol. VI (1935), p. 83.
- [5] WHITTAKER AND ROBINSON, *The Calculus of Observations*, London, 1924.

INSTITUTE OF ECONOMICS,  
UNIVERSITY OF OSLO,  
OSLO, NORWAY.

## NOTES

*This section is devoted to brief research and expository articles, notes on methodology and other short items.*

### A CRITERION FOR TESTING THE HYPOTHESIS THAT TWO SAMPLES ARE FROM THE SAME POPULATION

BY W. J. DIXON

**1. Introduction.** The purpose of this paper is to consider a criterion for testing the hypothesis that two samples have been drawn from populations with the same distribution function, assuming only that the cumulative distribution function common to the two populations is continuous. Let the two samples,  $O_n$  and  $O_m$ , be of size  $n$  and  $m$  respectively. We may assume  $n \leq m$  without loss of generality. Suppose the elements  $u_1, \dots, u_n$  of  $O_n$  are arranged in order from the smallest to the largest, that is,  $u_1 < u_2 < \dots < u_n$ . These may be represented as points along a line. The elements of  $O_m$  represented as points on the same line are then divided into  $(n + 1)$  groups by the first sample,  $O_n$ . Let  $m_1$  be the number of points having a value less than  $u_1$ ,  $m_i$  the number lying between  $u_i$  and  $u_{i+1}$ , ( $i = 1, 2, \dots, n$ ) and  $m_{n+1}$  the number greater than  $u_n$ , ( $m_{n+1} = m - m_1 - m_2 - \dots - m_n$ ). The criterion here proposed is<sup>1</sup>

$$(1) \quad C^2 = \sum_{i=1}^{n+1} \left( \frac{1}{n+1} - \frac{m_i}{m} \right)^2.$$

<sup>1</sup> A similar criterion

$$d^2 = \sum_{i=0}^n \left( \frac{i}{n} - \frac{\sum_{j=0}^i n_j}{n} \right)^2$$

for two samples of the same size was investigated (unpublished) by A. M. Mood. He found the mean and variance to be

$$E(d^2) = \frac{2n+1}{3n}, \quad \sigma_{d^2}^2 = \frac{8(n-1)(2n+1)}{45n^2}.$$

It can be seen that this is the sum of the squares of the differences between the ordinates of the two cumulative sample distributions calculated at the jumps of the first sample distribution.

**2. The mean and variance of  $C^2$ .** The only case of continuous cumulative distribution functions  $F(x)$  of any interest in statistics is that in which  $dF(x) = f(x) dx$ , where  $f(x)$  is a probability density function. Let us write:

$$p_1 = \int_{-\infty}^{u_1} f(x) dx, \quad p_2 = \int_{u_1}^{u_2} f(x) dx, \quad \dots, \quad p_{n+1} = \int_{u_n}^{\infty} f(x) dx,$$

where of course  $p_{n+1} = 1 - p_1 - p_2 - \dots - p_n$ .

Now, the joint distribution law of the  $p_i$  is

$$(2) \quad P(p_1, \dots, p_n) = n! dp_1 \dots dp_n$$

and the conditional distribution of the  $m_i$  given the  $p_i$  is

$$(3) \quad P(m_1, \dots, m_{n+1} | p_1, \dots, p_n) = \frac{n!}{m_1! \dots m_{n+1}!} p_1^{m_1} p_2^{m_2} \dots p_{n+1}^{m_{n+1}}.$$

Therefore the joint probability law of the  $m_i$  and  $p_i$  is

$$(4) \quad P(m, p) = \frac{n!n!}{m_1! \dots m_{n+1}!} p_1^{m_1} p_2^{m_2} \dots p_{n+1}^{m_{n+1}} dp_1 \dots dp_n.$$

Let  $\varphi(\theta) = \varphi(\theta_1, \dots, \theta_{n+1}) = E \left[ \exp \sum_{i=1}^{n+1} \theta_i \left( \frac{1}{n+1} - \frac{m_i}{n} \right) \right]$ ; then

$$(5) \quad E(C^2) = \sum_{i=1}^{n+1} \frac{\partial^2 \varphi}{\partial \theta_i^2} \Big|_{\theta=0},$$

$$(6) \quad E[(C^2)^2] = \sum_{i=1}^{n+1} \frac{\partial^4 \varphi}{\partial \theta_i^4} \Big|_{\theta=0} + \sum_{i \neq j} \frac{\partial^4 \varphi}{\partial \theta_i^2 \partial \theta_j^2} \Big|_{\theta=0}$$

and

$$(7) \quad \varphi(\theta) = \sum_m \int \exp \left[ \sum_{i=1}^{n+1} \theta_i \left( \frac{1}{n+1} - \frac{m_i}{n} \right) \right] P(m, p),$$

where  $\sum_m$  denotes the usual multinomial summation over all integral values of  $m_i \geq 0$  for which  $\sum m_i = m$  and the integration is over the generalized tetrahedron defined by  $p_i \geq 0$  and  $p_1 + p_2 + \dots + p_{n+1} \leq 1$ . If we perform the summation first, we obtain

$$(8) \quad \varphi(\theta) = n! e^{\sum_{i=1}^{n+1} \frac{\theta_i}{n+1}} \int (p_1 e^{-\frac{\theta_1}{n}} + \dots + p_{n+1} e^{-\frac{\theta_{n+1}}{n}})^m dp_1 \dots dp_n.$$

Differentiating twice with respect to  $\theta$ , and setting the  $\theta$ 's equal to zero, we get

$$\frac{\partial^2 \varphi}{\partial \theta_i^2} \Big|_{\theta=0} = n! \int \left[ \left( \frac{1}{n+1} \right)^2 + \left( \frac{1}{n} - \frac{2}{n+1} \right) p_i + \frac{m-1}{n} p_i^2 \right] dp_1 \dots dp_n.$$

If we now integrate and sum from one to  $n+1$ , we find

$$(9) \quad E(C^2) = \frac{n(n+m+1)}{m(n+1)(n+2)}.$$

Performing the operations indicated in (6), we obtain  $E[(C^2)^2]$  from which we subtract  $[E(C^2)]^2$  and have as the variance of  $C^2$ ,

$$\sigma_{C^2}^2 = \frac{4n(m-1)(m+n+1)(m+n+2)}{m^2(n+2)^2(n+3)(n+4)}.$$

**3. Significance values of  $C^2$ .** If we let  $C_\alpha^2$  be defined as the smallest value of  $C^2$  for which  $P(C^2 \geq C_\alpha^2) \leq \alpha$  then we can compute the value of  $C_\alpha^2$  fairly

TABLE I  
Values of  $C_\alpha^2$ .  $\alpha = 0.01, 0.05, 0.10$

$m \backslash n$	2	3	4	5	6	7	8	9	10
4	—	—	—						
	—	—	.800						
5	—	—	—	—					
	—	—	.800	.833					
	—	.750	.800	.833					
6	—	—	—	—	—				
	—	.750	.800	.833	.857				
	—	.750	.800	.556	.413				
7	—	—	—	—	—	—			
	—	.750	.800	.588	.612	.875			
	.667	.750	.555	.425	.449	.467			
8	—	—	.800	.833	.857	.656	.670		
	—	.750	.800	.594	.482	.469	.389		
	.667	.531	.425	.413	.357	.375	.358		
9	—	—	.800	.833	.660	.677	.543	.554	
	—	.750	.602	.448	.413	.431	.395	.381	
	.667	.552	.454	.389	.363	.356	.321	.307	
10	—	—	.800	.833	.677	.555	.549	.480	.449
	.667	.750	.480	.493	.437	.415	.349	.340	.349
	.487	.430	.380	.373	.357	.315	.309	.280	

readily for small values of  $m$  and  $n$ . The values of  $C^2$  for  $m, n \leq 10$  are given in Table I for  $\alpha = 0.01, 0.05$  and  $0.10$ . Since the distribution of  $C^2$  is not continuous the probabilities  $P(C^2 \geq C_\alpha^2)$  will, in general, be less than  $\alpha$ .



It will be seen that if  $m$  and  $n$  increase indefinitely in the ratio  $n/m = \gamma$ , then  $nC^2$  converges stochastically to  $\gamma + 1$  whereas  $nC^2$  ranges from 0 to  $n^2/(n+1)$  which indicates a tail to the right. This suggests that for larger values of  $m$  and  $n$ , it is reasonable to try to fit the distribution of  $nC^2$  by the method of moments using a distribution of the form

$$(11) \quad \frac{(kx^2)^{\frac{1}{2}\nu-1}}{2^{\frac{1}{2}}\Gamma(\frac{1}{2}\nu)} e^{-\frac{1}{2}kx^2} d(kx^2)$$

which has

$$E(x^2) = \frac{\nu}{k}, \quad \sigma_{x^2}^2 = \frac{2\nu}{k^2}.$$

Setting  $x^2 = nC^2$ , we see that we can consider  $nkC^2$  distributed as  $\chi^2$  with  $\nu$  degrees of freedom. Of course,  $\nu$  is not necessarily an integer, but  $\chi^2$  tables may be used for approximate values of the probability that  $nkC^2$  will exceed certain values,<sup>2</sup> or the values of  $nkC^2$  that will be exceeded a certain per cent of the time.<sup>3</sup> More exact values of these probabilities that  $nkC^2$  will exceed a certain value may be found from a table of the incomplete Gamma function.<sup>4</sup>

To calculate  $k$  and  $\nu$  directly, the following formulas obtained by equating the mean and variance of (11) to the mean and variance of  $nC^2$  may be used:

$$(12) \quad k = am(n+2)/n, \quad \nu = an(n+m+1)/(n+1),$$

where

$$a = \frac{m(n+3)(n+4)}{2(m-1)(m+n+2)(n+1)}.$$

If the fitted curve (11) is used to obtain significance values of  $nC^2$ , there is a tendency toward rejecting slightly over 100%, especially for small values of  $m$  and  $n$ . The error is probably due to fitting a curve having an infinite range. The discrepancy decreases as  $m$  and  $n$  increase.

The goodness of fit at the 0.01, 0.05 and 0.10 significance levels was tested for two cases.

Case 1.  $n = 9$ ,  $m = 10$ ;  $nk = \frac{2880}{11}$ ,  $\nu = \frac{52}{11}$ .

The exact distribution in the region under consideration is the following:

$C_0^2$	...	.26	.28	.30	.32	.34	.36	.40	.42	.44	.48	...
$P(C^2 \geq C_0^2)$	...	.121	.090	.082	.072	.037	.033	.025	.025	.015	.007	...

The values of  $C_\alpha^2$  from the fitted curve are  $C_{.01}^2 = 0.422$ ,  $C_{.05}^2 = 0.323$  and  $C_{.10}^2 = 0.277$ . The double rule indicates the divisions (from the fitted curve) for  $\alpha = 0.01, 0.05$  and  $0.10$ .

<sup>1</sup> Karl Pearson, *Tables for Statisticians and Biometricians*, part 1, Table XII.

<sup>2</sup> R. A. Fisher, *Statistical Methods for Research Workers*, Table III.

<sup>4</sup> *Tables of the Incomplete Gamma Function*, Biometrika Office, London.

Case 2.  $n = 12, m = 12; nk = 65.068, \nu = 8.938$ .

The important part of the exact distribution for our purposes is:

$C_0^2$	...	.215	.229	.243	.256	.270	...	.326	.340	.354	.381	...
$P(C^2 \geq C_0^2)$	...	.120	.109	.078	.057	.046	...	.017	.014	.011	.009	...

The values of  $C_\alpha^2$  from the fitted curve are  $C_{.01}^2 = 0.3315, C_{.05}^2 = 0.2587$  and  $C_{.10}^2 = 0.2244$ .

**4. Examples.** 1. Two samples of ten members each are drawn and it is desired to test, using a rejection region of size  $\alpha$ , the hypothesis that these two samples could have originated from the same population about which nothing is assumed except that it is continuous. The first sample was found to divide the second sample into the following groups: 0, 0, 0, 3, 0, 4, 0, 0, 2, 1, 0.

$$C^2 = (\bar{r}_1 - \bar{r}_0)^2 + (\bar{r}_1 - \bar{r}_0)^2 + (\bar{r}_1 - \bar{r}_0)^2 + (\bar{r}_1 - \bar{r}_0)^2 + 7(\bar{r}_1)^2 = .209$$

which we see from Table I is not a significant value even for  $\alpha = 0.10$  since  $C_{.10}^2 = 0.269$ .

2. A sample of 15 divides a second of 25 into the following 16 groups: 0, 1, 0, 0, 5, 4, 1, 3, 9, 0, 0, 1, 0, 1, 0, 0.

$$C^2 = (\bar{r}_6 - \bar{r}_5)^2 + (\bar{r}_6 - \bar{r}_5)^2 + (\bar{r}_6 - \bar{r}_5)^2 + (\bar{r}_6 - \bar{r}_5)^2 + 4(\bar{r}_6 - \bar{r}_5)^2 + 8(\bar{r}_6)^2$$

$$nC^2 = 2.302 \quad k = 7.511 \quad \nu = 10.19$$

$$nkC^2 = 17.295$$

which gives a significant value for  $\alpha = 0.10$  but not for  $\alpha = 0.05$ , since  $nkC_{.10}^2 = 16.233, nkC_{.05}^2 = 18.568$ . Actually  $P(nkC^2 > 17.29) = .077$ .

**5. Remarks.** If we set  $W$  equal to the number of  $m_i$  which are zero and  $V = n + 1 - W$  then  $V$  is the number of non-zero  $m_i$ ; further,  $2V \cong U$  where  $U$  is the total number of runs, the criterion proposed in the paper of Wald and Wolfowitz in the present issue of the *Annals of Mathematical Statistics*. Now,

$$(13) \quad W = \lim_{x_1, \dots, x_{n+1} \rightarrow 0} \sum_{i=1}^{n+1} x_i^{m_i},$$

so that, setting

$$(14) \quad \Phi = \sum_m \int \exp \left[ \sum_{i=1}^{n+1} \theta_i \left( \frac{1}{n+1} - \frac{m_i}{m} \right) \right] \sum_{i=1}^{n+1} x_i^{m_i} P(m, p),$$

analogous to (7), we have

$$E(WC^2) = \lim_{x_1, \dots, x_{n+1} \rightarrow 0} \sum_{k=1}^{n+1} \frac{\partial^2 \Phi}{\partial \theta_k^2} \Big|_{\theta=0}$$

from which we can find

$$\rho_{vc^2} \sigma_v \sigma_{c^2} = \frac{2n(1-m)}{m(n+2)(m+n)}$$

and

$$\rho_{vc^2}^2 \cong \rho_{vc^2}^2 - \frac{(n+3)(n+4)(m+n-1)}{(n+1)(m+n+1)(m+n+2)}.$$

If  $n/m = \gamma$  (a fixed constant) and  $n$  is large

$$\rho^2 \cong \frac{n}{n+m}.$$

$\rho^2$  will be near 1 when  $n$  is much larger than  $m$ . This corresponds, in computing  $C^2$ , to dividing the smaller sample into subgroups by the larger. In this case  $U$  and  $C^2$  give essentially the same information. When  $m$  and  $n$  are more nearly equal the two criteria are quite different. For  $n > m$ ,  $C^2$  has fewer possible values than for  $n < m$ , and is therefore a more sensitive test when  $n < m$ .

While it is doubtful that this test is biased for large samples, this question will not be considered in the present note.

PRINCETON UNIVERSITY,  
PRINCETON, N. J.

## SIGNIFICANCE TEST FOR SPHERICITY OF A NORMAL $n$ -VARIATE DISTRIBUTION

BY JOHN W. MAUCHLY

**1. Introduction.** This note is concerned with testing the hypothesis that a sample from a normal  $n$ -variate population is in fact from a population for which the variances are all equal and the correlations are all zero. A population having this symmetry will be called "spherical." Under a linear orthogonal transformation of variates, a spherical population remains spherical, and consequently the features of a sample which furnish information relevant to this hypothesis must be invariant under such transformations.

A situation for which this test is indicated arises when the sample consists of  $N$   $n$ -dimensional vectors, for which the variates are the  $n$  components along coordinate axes known to be mutually perpendicular, but having an orientation which is, a priori at least, quite arbitrary. A specific application for two dimensions, treated elsewhere [1], may be mentioned. Each of  $N$  days furnishes a sine and a cosine Fourier coefficient for a given periodicity, and these, when plotted as ordinate and abscissa, yield a somewhat elliptical cloud of  $N$  points. The sine and cosine functions are orthogonal, and their variances have

equal expectancies for a random series. The arbitrary nature of the orientation of axes appears here as the arbitrary choice of phase, or origin of time. Of the five ellipses studied, three could easily have come from circular populations (random), and two showed highly significant ellipticity.

**2. Likelihood ratio criterion for sphericity.** The method of Neyman and Pearson [2] will be used to derive a test criterion which seems entirely suitable. Let  $\Omega$  be the class of all normal  $n$ -variate populations, and let  $\omega$  be the subclass of all normal  $n$ -variate populations satisfying the hypothesis of "sphericity." The likelihood ratio criterion is obtained by taking the ratio of the maximum of the likelihood for variation of all population parameters specifying  $\omega$ , to the maximum of the likelihood for variation of all population parameters specifying  $\Omega$ . That is,

$$(1) \quad \lambda = \frac{P(\omega \text{ max})}{P(\Omega \text{ max})}.$$

For the set  $\Omega$ , the probability law for a single observation of the  $n$  variates may be written:

$$(2) \quad P = K |a_{ij}|^{-1} e^{-\frac{1}{2} \sum_{i,j} c_{ij} (x_i - a_i)(x_j - a_j)} \quad (i, j = 1, 2 \dots n),$$

where  $c_{ij}$  is an element of the matrix  $||a_{ij}||^{-1}$ , the  $a_{ij}$  being variances and covariances,  $a_i$  is the mean value of the variate  $x_i$  in the population, and  $K$  is a constant the value of which does not concern us here. Then a sample of  $N$  from  $\Omega$  has the probability,

$$(3) \quad P = K^N |a_{ij}|^{-1N} e^{-\frac{1}{2} \sum_{i,j} c_{ij} \sum_{\alpha=1}^N (x_{i\alpha} - a_i)(x_{j\alpha} - a_j)}.$$

Letting

$$(4) \quad \sum_{\alpha=1}^N x_{i\alpha} = N\bar{x}_i \quad \text{and} \quad \sum_{\alpha=1}^N (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) = N s_{ij},$$

differentiating the logarithm of  $P$  with respect to the parameters  $a_i$  and  $a_{ij}$ , and setting these derivatives equal to zero, the maximum likelihood estimates,

$$(5) \quad \hat{a}_i = \bar{x}_i; \quad \hat{a}_{ij} = s_{ij},$$

are obtained. Substituting these values in equation (3) we find that the maximum value of the likelihood is

$$(6) \quad P(\Omega \text{ max}) = K^N |s_{ij}|^{-1N} e^{-1N}.$$

The derivation of  $P(\omega \text{ max})$  proceeds upon similar lines, but is simpler, for the probability law for the set  $\omega$  is obtained from (3) by setting

$$(7) \quad c_{ij} = c\delta_{ij},$$

where  $c$  is any positive constant, and  $\delta_{ij} = 0$  if  $i \neq j$  and 1 if  $i = j$ . The result is found to be

$$(8) \quad P(\omega \max) = K^N (s_0)^{-\frac{1}{2}Nn} e^{-\frac{1}{2}N}$$

where  $s_0$  is defined by

$$(9) \quad ns_0 = \sum_{i=1}^n s_{ii}.$$

The likelihood ratio criterion is therefore

$$(10) \quad \lambda_s = \left[ \frac{|s_{ij}|^{\frac{1}{2}}}{(s_0)^{\frac{1}{2}n}} \right]^N.$$

It will be convenient to designate the  $N$ th root of this statistic as  $L_{sn}$ , where the second subscript indicates the number of variates:

$$(11) \quad L_{sn} = \frac{|s_{ij}|^{\frac{1}{2}}}{s_0^{\frac{1}{2}n}}.$$

**3. The moments of the distribution of  $L_{sn}$  when the population is spherical.** The distribution of  $L_{sn}$  cannot be easily obtained in explicit form for a general  $n$ , but the moments of  $L_{sn}$  when the hypothesis tested is true are easily found.

Note first that  $L_{sn}$  may be resolved into two factors which are, when the population is spherical, statistically independent:

$$(12) \quad L_{sn} = \frac{(s_1 s_2 s_3 \cdots s_n)^{\frac{1}{2}}}{s_0^{\frac{1}{2}n}} \cdot |r_{ij}|^{\frac{1}{2}}.$$

The first factor is just the one appropriate for testing the equality of the  $n$  variances when the orientation of the coordinate axes is fixed in advance, while the second factor is the square root of the determinant of correlation coefficients. The moments of the distributions of these two statistics are known [3], and since the two are independent (for zero correlation in the population), we may write:

$$(13) \quad M_h(L_{sn}) = M_h(A)M_h(B),$$

where  $A$  and  $B$  are used to indicate the two factors, and  $M_h$  indicates the  $h$ th moment. The moments are given by

$$(14) \quad M_h(L_{sn}) = \prod_{i=1}^n \left[ \frac{\Gamma(\frac{1}{2}(N-i+h))}{\Gamma(\frac{1}{2}(N-i))} \right] n^{\frac{1}{2}nh} \frac{\Gamma(\frac{1}{2}(n(N-1)))}{\Gamma(\frac{1}{2}(n(N-1+h)))}.$$

**4. Significance test for  $n = 2$ .** For  $n = 1$ ,  $M_h(L_{s1}) = 1$  for any  $h$ , as it should, since  $L_{s1}$  is then identically 1, and the concept of sphericity is meaningless. For  $n = 2$ , the expression (14) reduces to,

$$(15) \quad M_h(L_{s2}) = \frac{\Gamma(N-2+h)\Gamma(N-1)}{\Gamma(N-1+h)\Gamma(N-2)} = \frac{N-2}{N-2+h}.$$

and the distribution is thus found to be

$$(16) \quad D(L_{n2}) = (N - 2)L_{n2}^{N-3} dL_{n2}.$$

Thus for  $n = 2$ , the significance of the value of  $L_{n2}$  obtained from a given sample of  $N$  points in a plane is simply

$$(17) \quad P(L_{n2} < L'_{n2}) = L'_{n2}{}^{N-2}.$$

These results for  $n = 2$  were obtained by another method in [1].

**5. Significance test for  $n = 3$ .** For  $n = 3$  and higher values of  $n$ , no simple expression for the distribution seems obtainable. In this case it appears reasonable to fit a Pearson curve of the type,

$$(18) \quad y = Kx^{p-1}(1 - x)^{q-1},$$

by adjusting  $p$  and  $q$  so as to obtain agreement with the first two moments of the actual distribution. The calculations were carried out for  $L_{n3}^2$  rather than  $L_{n3}$  itself, to simplify the moment expressions. The first moment of  $L_{n3}^2$  is the second moment of  $L_{n3}$ , and is given as a function of  $N$  by the equation,

$$(19) \quad \mu_1(N) = \frac{(3N - 6)(3N - 9)}{(3N - 2)(3N - 1)}.$$

Recurrence relations, similar to those noted by Lengyel [4] in carrying out a similar task, hold for the moments of  $L_{n3}^2$ ; hence,

$$(20) \quad \mu_2(N) = \mu_1(N)\mu_1(N + 2).$$

Explicit solution of the equations for  $p$  and  $q$  in terms of  $N$  is possible:

$$(21) \quad p = \frac{(9N + 5)(N - 2)(N - 3)}{2(9N^2 - 8N - 15)}$$

$$(22) \quad q = \frac{2(9N - 13)(9N + 5)}{9(9N^2 - 8N - 15)}.$$

For values of  $N > 30$ , acceptable approximations to  $p$  and  $q$  are obtained by carrying out the division indicated in (21) and (22):

$$(23) \quad p = \frac{1}{2}(N - 4) + 2/9 + 70/81(N + 1) \dots,$$

$$(24) \quad q = 2 + \frac{140}{9(3N - 2)^2} \dots$$

The values of  $p$  and  $q$  are given in Table I so that those desiring other than the standard significance levels may readily enter the Pearson tables.

For  $N$  a multiple of 4 from 8 to 48, and a multiple of 10 from 50 to 100, the significance levels were taken from the Incomplete Beta-Function Tables, using adequate interpolation. The final Table I was then prepared by filling in the skeleton table by interpolation with respect to  $N$ .

From the results of Wilks [5] it follows that  $-2N \log_e L_{n3}$  is, for large  $N$ ,

TABLE I

5%, 1%, and 0.1% levels of significance for the 3-dimensional sphericity criterion,  $L_{13}^2 = \lambda^{2/N}$ , and the values of  $p$  and  $q$  for the Pearson Type I curves used in calculating these levels

$N$	5%	1%	0.1%	$p$	$q$
8	0.172	0.083	0.030	2.3239	2.0312
10	.278	.165	.080	3.3044	2.0194
12	.366	.243	.139	4.2911	2.0131
14	.436	.312	.197	5.2816	2.0095
16	.494	.372	.252	6.2744	2.0072
18	.541	.423	.301	7.2688	2.0057
20	.580	.466	.346	8.2642	2.0046
22	.614	.504	.386	9.2605	2.0038
24	.642	.538	.422	10.2574	2.0032
26	.667	.567	.454	11.2548	2.0027
28	.689	.593	.483	12.2526	2.0023
30	.708	.616	.510	13.2506	2.0020
32	.724	.637	.534	14.2488	2.0018
34	.739	.655	.555	15.2473	2.0016
36	.753	.672	.575	16.2458	2.0014
38	.765	.687	.594	17.2447	2.0012
40	.776	.701	.610	18.2435	2.0011
42	.786	.714	.626	19.2425	2.0010
44	.795	.726	.640	20.2416	2.0009
46	.804	.736	.653	21.2408	2.0008
48	.811	.746	.665	22.2400	2.0008
50	.819	.756	.677	23.2394	2.0007
55	.834	.776	.703	*	*
60	.848	.793	.725	28.2365	2.0005
65	.859	.808	.744	*	*
70	.869	.821	.760	33.2345	2.0004
75	.877	.832	.775	*	*
80	.885	.842	.788	38.2328	2.0003
85	.891	.851	.799	*	*
90	.897	.859	.809	43.2317	2.0002
95	.902	.866	.819	*	*
100	.907	.872	.827	48.2308	2.0002

\*No values for  $p$  and  $q$  were calculated for these values of  $N$ ; the levels were obtained by interpolation (see text).

distributed approximately like  $\chi^2$  with  $n(n-1)/2$  degrees of freedom. However, equation (24) above suggests that for large  $N$  one may get a very good

approximation (for  $n = 3$ ) by setting  $q = 2$ ; the significance test for  $n = 3$  then becomes,

$$(25) \quad P(L_{.3} < L'_{.3}) = \frac{1}{2} L'_{.3}{}^{N-4} [(N-2) - (N-4)L'_{.3}].$$

Probably similar approximations can be found for other values of  $n$ . It is a pleasure to acknowledge the helpful comments and advice which I received from Mr. A. M. Mood of Princeton. Recognition is also due Mr. Wallace Broy, a student assistant under the National Youth Administration, who aided in the computations.

#### REFERENCES

- [1] J. W. MAUCHLY, *Terr. Magn.*, Vol. 45 (1940) (In press).
- [2] NEYMAN AND PEARSON, *Trans. Roy. Soc.*, A, 231 (1933), p. 295.
- [3] S. S. WILKS, *Biometrika*, Vol. 24 (1932), p. 471.
- [4] B. A. LENGUEL, *Annals of Math. Stat.*, Vol. 10 (1939), p. 365.
- [5] S. S. WILKS, *Annals of Math. Stat.*, Vol. 9 (1938), p. 60.

URSINUS COLLEGE,  
COLLEGEVILLE, PA.

### A SIMPLE SAMPLING EXPERIMENT ON CONFIDENCE INTERVALS

By S. KULLBACK AND A. FRANKEL

**1. Introduction.** In order to illustrate some of the notions of the theory of confidence or fiducial limits in connection with a course in Statistical Inference at the George Washington University, we had the class carry out certain simple experiments, following a suggestion in one of Neyman's papers on Statistical Estimation [1]. In the belief that the experimental data may be of interest to others, we present the results herein.

**2. The problem.** We consider the problem of estimating the range  $\theta$  of a rectangular population defined by  $p(x, \theta) dx = dx/\theta$ ,  $0 \leq x \leq \theta$  and in particular, for simplicity, we limit ourselves to samples of two and four. We consider three possible approaches to the problem, viz., by using (a) the sample range (b) the sample average or total (c) the larger (largest) sample value. Let us consider each in turn.

(a) *Sample range.* Wilks [2] has shown that for samples of  $n$  and confidence coefficient  $1 - \alpha$ , the confidence or fiducial limits for the population range  $\theta$  are given by  $r$  and  $r/\psi_\alpha$ , where  $r$  is the sample range and  $\psi_\alpha$  is determined by

$$(1) \quad \psi_\alpha^{n-1} [n - (n-1)\psi_\alpha] = \alpha.$$

For  $n = 2$ ,  $\alpha = 0.19$  and  $n = 4$ ,  $\alpha = 0.1792$ , (1) yields  $\psi_\alpha = 0.1$  and  $\psi_\alpha = 0.4$  respectively. Accordingly, for samples of two with confidence coefficient



$1 - \alpha = 0.81$ , and for samples of four with confidence coefficient  $1 - \alpha = 0.8208$ , the confidence interval is respectively given by

$$(2) \quad (r, 10r) \quad \text{and} \quad (r, 2.5r).$$

The length,  $\lambda_r$ , of the confidence interval is respectively  $9r$  and  $1.5r$ . Using the distribution of  $r$ ,  $n(n-1)(\theta-r)r^{n-2}/\theta^n$ , we have for samples of two:  $E(\lambda_r) = 3\theta$ ,  $\sigma_{\lambda_r} = 2.1213\theta$ , and for samples of four:  $E(\lambda_r) = 0.9\theta$ ,  $\sigma_{\lambda_r} = 0.3\theta$ .

(b) *Sample total.* Following Neyman [1, p. 357] let us denote by  $A(\theta)$  the region defined by

$$(3) \quad \theta - \Delta \leq x_1 + x_2 \leq \theta + \Delta$$

where  $\theta$  is the population range,  $x_1$  and  $x_2$  the sample values of the sample  $E_2$  and  $\Delta$  is selected so as to have  $P\{E_2 \in A(\theta) \mid \theta\} = 1 - \alpha$ . It is readily found that  $P\{E_2 \in A(\theta) \mid \theta\} = [\theta^2 - (\theta - \Delta)^2]/\theta^2 = 1 - \alpha$  from which we find that  $\Delta = \theta(1 - \alpha^{1/2})$ . Accordingly (3) becomes  $\theta\alpha^{1/2} \leq x_1 + x_2 \leq \theta(2 - \alpha^{1/2})$ , yielding the confidence limits  $(x_1 + x_2)/(2 - \alpha^{1/2})$ ,  $(x_1 + x_2)/\alpha^{1/2}$ . For the confidence coefficient  $1 - \alpha = 0.81$  the confidence interval is given by

$$(4) \quad [0.6394(x_1 + x_2), 2.2941(x_1 + x_2)].$$

The length of the confidence interval is given by  $\lambda_T = 1.6547(x_1 + x_2)$  so that  $E(\lambda_T) = 1.6547\theta$ ,  $\sigma_{\lambda_T} = 0.6755\theta$ .

Let us denote by  $A'(\theta)$  the region defined by

$$(5) \quad 2\theta - \Delta \leq x_1 + x_2 + x_3 + x_4 \leq 2\theta + \Delta,$$

where  $\theta$  is the population range,  $x_1, x_2, x_3, x_4$  the sample values of the sample  $E_4$  and  $\Delta$  is selected so as to have  $P\{E_4 \in A'(\theta) \mid \theta\} = 1 - \alpha$ . Using the known distribution of the sample average [3] and  $1 - \alpha = 0.8208$ , it is readily found that

$$P\{E_4 \in A'(\theta) \mid \theta\} = \frac{16}{3} \left\{ \frac{\Delta}{4\theta} - 8 \left( \frac{\Delta}{4\theta} \right)^3 + 12 \left( \frac{\Delta}{4\theta} \right)^4 \right\} = 0.8208$$

from which we find that  $\Delta = 0.788\theta$ . Accordingly, (5) becomes  $1.212\theta \leq x_1 + x_2 + x_3 + x_4 \leq 2.788\theta$ , yielding the confidence interval

$$(6) \quad [0.3587(x_1 + x_2 + x_3 + x_4), 0.8251(x_1 + x_2 + x_3 + x_4)].$$

The length of the confidence interval is given by  $\lambda_T = 0.4664(x_1 + x_2 + x_3 + x_4)$  so that  $E(\lambda_T) = 0.9328\theta$  and  $\sigma_{\lambda_T} = 0.2679\theta$ .

(c) *Larger (largest) sample value.* Again following Neyman [1, p. 359] let us denote by  $A_1(\theta)$  the region defined by

$$(7) \quad q\theta \leq L \leq \theta$$

where  $\theta$  is the population range,  $L$  the larger of the two sample values  $x_1$  and  $x_2$  and  $q$ , a number between zero and unity, to be determined by  $P\{E_2 \in A_1(\theta) \mid \theta\} = 1 - \alpha$ . It is readily found that  $P\{E_2 \in A_1(\theta) \mid \theta\} = (\theta^2 - q^2\theta^2)/\theta^2 = 1 - \alpha$ ,

from which we find that  $q = \alpha^{1/2}$ . Accordingly, (7) becomes  $\theta\alpha^{1/2} \leq L \leq \theta$  yielding the confidence limits  $L, L/\alpha^{1/2}$ . For the confidence coefficient  $1 - \alpha = 0.81$  the confidence interval is given by

$$(8) \quad (L, 2.2941L).$$

TABLE I

No. of cases of coverage per set of 100 samples	Frequency					
	Range		Sum		Larger (Largest)	
$x$	Samples of two	Samples of four	Samples of two	Samples of four	Samples of two	Samples of four
69					1	
70						
71					1	
72						
73						1
74		1			1	
75						
76	4		3		4	1
77	2		6	1	2	
78	3		6		3	1
79	9	2	4	2	3	
80	3	1	6		4	
81	2	2	1		3	
82	2	1	6	1	2	5
83	3	3	3	1	5	3
84	3	2		1	4	1
85	3			3	2	
86	2	2		2	2	1
87	1	1	2	1		1
88			1	2	1	1
89	1		1	1		
90						
91	1					
	39	15	39	15	39	15
Average....	81.1	82.1	80.2	84.2	80.2	82.1

The length of the confidence interval is given by  $\lambda_L = 1.2941L$  so that using the distribution of  $L, nL^{n-1}dL$ , we have  $E(\lambda_L) = 0.8627\theta$  and  $\sigma_{\lambda_L} = 0.3050\theta$ .

Incidentally, since  $L \leq x_1 + x_2$  we have  $1.2941L < 1.6547(x_1 + x_2)$  so that

in every case, for samples of two, the confidence interval of procedure (c) is shorter than the confidence interval of procedure (b).

For samples of four, we consider the region (7) where  $L$  is the largest of the sample values  $x_1, x_2, x_3$  and  $x_4$  of the sample  $E_4$ . It is readily found that  $P\{E_4 \in A_1(\theta) \mid \theta\} = (\theta^4 - q^4\theta^4)/\theta^4 = 1 - \alpha$ , from which we find that  $q^4 = \alpha$ . For  $\alpha = 0.1792$ ,  $q = 0.6506$  so that (7) becomes  $0.6506\theta \leq L \leq \theta$  yielding the confidence interval

$$(9) \quad (L, 1.5370L).$$

The length of the confidence interval is given by  $\lambda_L = 0.5370L$  so that  $E(\lambda_L) = 0.4296\theta$  and  $\sigma_{\lambda_L} = 0.0877\theta$ .

TABLE II

	Sample size	Range		Sum		Larger (Largest)	
		Theoretical	Observed	Theoretical	Observed	Theoretical	Observed
Confidence Coefficient	2	.8100	.8110	.8100	.802	.8100	.8020
	4	.8208	.8210	.8208	.842	.8208	.8210
Average length of confidence interval per set of 100 samples	2	3.0000	2.9660	1.6547	1.6441	.8627	.8556
	4	.9000	.8976	.9328	.9296	.4296	.4272
Standard deviation of average length of confidence interval	2	.2121	.2133	.0676	.0581	.0305	.0293
	4	.0300	.0335	.0268	.0140	.0088	.0093

**3. The Experimental Data.** We considered the rectangular population with  $\theta = 1$  and obtained the sample values by using pairs of digits obtained from Tippett's random sample tables [4]. Using these observed values the confidence intervals given by (2), (4), (6), (8) and (9) were computed and the number of cases in which the value  $\theta = 1$  was covered, noted. In all, 3900 samples of two were observed, subdivided into 39 sets of 100 each. The samples of four were obtained by combining pairs of samples of two and there were studied 1500 samples of four, subdivided into 15 sets of 100 each. Table I gives the observed distribution of the number of cases of coverage per set of 100 samples of two and of four. The length of the confidence interval obtained by each of the three procedures was obtained and the observed mean and standard deviation of the distribution of the average length of the confidence interval per set of 100 samples computed. (Since they are averages of 100 values, these observations are practically normally distributed.) Table II summarizes these results.

## REFERENCES

- [1] J. NEYMAN, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability," *Phil. Trans. Roy. Soc. Series A* Vol. 236 (1937), pp. 333-380.
- [2] S. S. WILKS, "Fiducial Distributions in Fiducial Inference," *Annals of Math. Stat.* Vol. IX (1938), pp. 272-280.
- [3] P. HALL, "The distribution of means for samples of size  $N$  drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable," *Biometrika*, Vol. 19 (1927), pp. 240-244.
- [4] L. H. C. TIPPETT, *Tracts for computers No. 16*, Camb. Univ. Press, 1927.

THE GEORGE WASHINGTON UNIVERSITY,  
WASHINGTON, D. C.

## THE NUMERICAL COMPUTATION OF THE PRODUCT OF CONJUGATE IMAGINARY GAMMA FUNCTIONS

BY A. C. COHEN, JR.

The difference equation

$$(1) \quad \begin{aligned} f_{z+1} &= x^2 + c_1x + c_2 \\ f_z &= x^2 + c_3x + c_4 \end{aligned}$$

was used by Professor Harry C. Carver [1] as the basis for graduating frequency distributions in a manner analogous to the use of the differential equation

$$\frac{1}{y} \frac{dy}{dx} = \frac{a - x}{b_0 + b_1x + b_2x^2}$$

in the Pearson system of frequency curves. In order to determine a particular  $f_z$  by Professor Carver's method it was necessary to perform the complete graduation from the lower limit of the range up to and including the required  $f_z$ . When  $x$  is large and only isolated values of  $f_z$  are required it seems desirable to have a method for computing  $f_z$  directly, and the present note seeks to accomplish this purpose.

It is well known [2] that the difference equation

$$(2) \quad \frac{f_{z+1}}{f_z} = a \frac{(x - \alpha_1)(x - \alpha_2) \dots (x - \alpha_n)}{(x - \beta_1)(x - \beta_2) \dots (x - \beta_m)}$$

has the solution

$$(3) \quad f_z = w_z a^z \frac{\Gamma(x - \alpha_1) \dots \Gamma(x - \alpha_n)}{\Gamma(x - \beta_1) \dots \Gamma(x - \beta_m)},$$

where  $w_z$  is a periodic function of  $x$  ( $w_z = w_{z+n} = \dots = k$ ) and  $\Gamma(x + 1)$  for  $x$ , a positive real number may be defined in the usual manner by the second Euler integral

$$(4) \quad \Gamma(x + 1) = \int_0^\infty t^x e^{-t} dt$$

which obeys the recursion formula

$$(5) \quad \Gamma(x+1) = x\Gamma(x).$$

When  $x$  is a positive integer

$$(6) \quad \Gamma(x+1) = x!$$

Equation (1) is seen to be a special case of (2) for  $n = m = 2$  and accordingly, the solution may be written as

$$(7) \quad f_s = K \frac{\Gamma(x - \alpha_1)\Gamma(x - \alpha_2)}{\Gamma(x - \beta_1)\Gamma(x - \beta_2)},$$

where  $\alpha_1$  and  $\alpha_2$  are roots of  $x^2 + c_1x + c_2 = 0$  and  $\beta_1$  and  $\beta_2$  are roots of  $x^2 + c_3x + c_4 = 0$ . The following simple examples illustrate three special cases of this solution.

I. All  $\alpha$ 's and  $\beta$ 's are integers.

$$\frac{f_{s+1}}{f_s} = \frac{2(x^2 + 9x + 20)}{x^2 + 5x + 6}$$

has the solution

$$f_s = K2^s \frac{\Gamma(x+4)\Gamma(x+5)}{\Gamma(x+2)\Gamma(x+3)}$$

which, with the aid of recursion formula (5) can readily be verified by direct substitution.

II. Either the  $\alpha$ 's and/or the  $\beta$ 's are real irrational numbers.

$$\frac{f_{s+1}}{f_s} = \frac{x^2 + 5x + 6}{x^2 + 3x + 1}$$

has the solution

$$f_s = K \frac{\Gamma(x+2)\Gamma(x+3)}{\Gamma[x + \frac{1}{2}(3 - \sqrt{5})]\Gamma[x + \frac{1}{2}(3 + \sqrt{5})]}$$

which, with the aid of the recursion formula (5) can also be verified by direct substitution.

III. Either the  $\alpha$ 's and/or the  $\beta$ 's are complex.

$$\frac{f_{s+1}}{f_s} = \frac{x^2 + 8x + 17}{x^2 + 10x + 29}$$

has the solution

$$f_s = K \frac{\Gamma(x+4+i)\Gamma(x+4-i)}{\Gamma(x+5+2i)\Gamma(x+5-2i)}.$$

Since the recursion formula (5) is also valid for complex arguments [3], this solution can be verified by direct substitution just as in the first two cases.

The evaluation of  $f_s$  for a given  $x$  in cases I and II involves only computation

of quantities of the form  $\Gamma(x)$  which can be accomplished through the use of existing tables of Gamma Functions for small values of  $x$  and through application of Stirling's formula for large values of  $x$ . Evaluation of  $f_s$  in case III, however, involves the computation of quantities of the form  $\Gamma(u + iv)\Gamma(u - iv)$ , a problem which seems to have escaped previous attention. The remainder of the present discussion will center about this quantity.

The Gamma Function for a real positive argument has been defined by equation (4), but for the present purposes, it is more expedient to use the definition

$$(8) \quad \Gamma(z) = \lim_{n \rightarrow \infty} \frac{n!n^z}{z(z+1) \cdots (z+n)}$$

which is valid for all values of the complex argument  $z$  except at the poles ( $z = -1$ ;  $z = -2$ , etc.). The above definition is equivalent to (4) at all points where (4) is valid [3].

From equation (8), it immediately follows that  $\Gamma(u + iv)\Gamma(u - iv)$  is a real number. In fact, we have

$$\Gamma(u + iv)\Gamma(u - iv) = \lim_{n \rightarrow \infty} \frac{(n!)^2 n^{2u}}{[u^2 + v^2][(u+1)^2 + v^2] \cdots [(u+n)^2 + v^2]}.$$

We now develop a formula applicable in evaluating this quantity when  $u$  is a sufficiently small positive integer. As a consequence of equation (8) it can be shown that [3]

$$(9) \quad \Gamma(z)\Gamma(1-z) = \frac{\sin \pi z}{\pi}$$

Let  $z = iv$  in the above equation and we immediately obtain the result

$$(10) \quad \Gamma(iv)\Gamma(-iv) = \frac{2\pi v^{-1}}{e^{\pi v} - e^{-\pi v}}.$$

When  $u$  is a positive integer, we may write

$$(11) \quad \Gamma(u + iv) = (u - 1 + iv)(u - 2 + iv) \cdots (iv)\Gamma(iv),$$

$$(12) \quad \Gamma(u - iv) = (u - 1 - iv)(u - 2 - iv) \cdots (-iv)\Gamma(-iv).$$

The product of (11) by (12) gives

$$\Gamma(u + iv)\Gamma(u - iv) = v^2(v^2 + 1) \cdots (v^2 + u - 1^2)\Gamma(iv)\Gamma(-iv)$$

which upon substitution of the value found in Equation (10) for  $\Gamma(iv)\Gamma(-iv)$  becomes

$$(13) \quad \Gamma(u + iv)\Gamma(u - iv) = \frac{2\pi v}{e^{\pi v} - e^{-\pi v}} \prod_{r=1}^{u-1} (v^2 + r^2).$$

To obtain a result that is applicable when  $u$  is not a positive integer, we make use of Stirling's formula for complex arguments. Lipschitz [4] proves

$$(14) \quad \begin{aligned} \log \Gamma(z) &= \log \sqrt{2\pi} + (z - \frac{1}{2}) \log z \\ &\quad - z + (-1)^m \sum_{n=0}^m \frac{B_{2n+1}}{(2n+1)(2n+2)} \frac{1}{z^{2n+1}} \end{aligned}$$

and that the remainder after the  $m$ th term is

$$v_{m+1} = \frac{(-1)^{m+1} B_{2m+3}}{(2m+3)(2m+4)} \frac{1}{z^{2m+3}} (\epsilon + \epsilon' i),$$

where  $\epsilon < 1$ ;  $\epsilon' < 1$ .  $B_{2m+1}$  designates the Bernoulli numbers. ( $B_1 = \frac{1}{2}$ ;  $B_3 = \frac{1}{6}$ ;  $B_5 = \frac{1}{42}$ ; etc.) We are thus able to write

$$(15) \quad \begin{aligned} \log \Gamma(u + iv) &= \log \Gamma(Re^{i\varphi}) \\ &= \log \sqrt{2\pi} + (Re^{i\varphi} - \frac{1}{2})(\log R + i\varphi) \\ &\quad - Re^{i\varphi} + \sum_{n=0}^{\infty} \frac{(-1)^n B_{2n+1}}{(2n+1)(2n+2)} \frac{e^{-(2n+1)i\varphi}}{R^{2n+1}}, \end{aligned}$$

where  $\varphi = \tan^{-1} \frac{v}{u}$  and  $R = \sqrt{u^2 + v^2}$ ;

$$(16) \quad \begin{aligned} \log \Gamma(u - iv) &= \log \Gamma(Re^{-i\varphi}) \\ &= \log \sqrt{2\pi} + (Re^{-i\varphi} - \frac{1}{2})(\log R - i\varphi) \\ &\quad - Re^{-i\varphi} + \sum_{n=0}^{\infty} \frac{(-1)^n B_{2n+1}}{(2n+1)(2n+2)} \frac{e^{(2n+1)i\varphi}}{R^{2n+1}}. \end{aligned}$$

Adding (15) and (16), we obtain

$$\begin{aligned} \log \Gamma(u + iv) \Gamma(u - iv) &= \log 2\pi + (e^{i\varphi} + e^{-i\varphi})R \log R - \log R \\ &\quad + Ri\varphi(e^{i\varphi} - e^{-i\varphi}) - R(e^{i\varphi} + e^{-i\varphi}) \\ &\quad + \sum_{n=0}^{\infty} \frac{(-1)^n B_{2n+1}}{(2n+1)(2n+2)} (e^{(2n+1)i\varphi} + e^{-(2n+1)i\varphi}) \frac{1}{R^{2n+1}} \end{aligned}$$

which upon being simplified becomes

$$(17) \quad \begin{aligned} \log \Gamma(u + iv) \Gamma(u - iv) \\ = \log 2\pi + (2u - 1) \log R - 2(\varphi v + u) + 2\psi(R, \varphi), \end{aligned}$$

where

$$(18) \quad \psi(R, \varphi) = \sum_{n=0}^{\infty} \frac{(-1)^n B_{2n+1}}{(2n+1)(2n+2)} \frac{1}{R^{2n+1}} \cos (2n+1)\varphi.$$

This result is somewhat similar to that obtained by Karl Pearson [5] in connection with the evaluation of the  $G_{(r)}$  integrals of his Type IV frequency

curve. If  $R > 1$ , the expansion of  $\psi(R, \varphi)$  is asymptotic and the greatest numerical value that the  $m$ th term can have is

$$\frac{B_{2m+1}}{(2m+1)(2m+2)} \cdot \frac{1}{R^{2m+1}}.$$

Thus according to Lipschitz results, the error committed in dropping all terms after the  $m$ th will not exceed:  $\pm \frac{B_{2m+1}}{(2m+1)(2m+2)} \frac{1}{R^{2m+1}}$ . The following table gives an indication of the size of the error:

Terms omitted after	Error committed in $\psi(R, \varphi)$ less than
1st	$\pm .0833 \ 3333/R$
2nd	$\pm .0027 \ 7777/R^3$
3rd	$\pm .0007 \ 9365/R^5$
4th	$\pm .0005 \ 9524/R^7$
5th	$\pm .0008 \ 4175/R^9$ .

It is now obvious that formula (18) will give satisfactory results whenever  $R$  is sufficiently large. The degree of accuracy required together with the value of  $R$  will determine the number of terms of  $\psi(R, \varphi)$  to be computed.

We now turn to the solution of the example under Case III and proceed to calculate  $f_4$ ,  $f_{15}$ , and  $f_{150}$  when  $f_0 = 29$ . We may write

$$K = 29 \frac{\Gamma(5+2i)\Gamma(5-2i)}{\Gamma(4+i)\Gamma(4-i)}.$$

Application of formula (13) gives

$$\Gamma(5+2i)\Gamma(5-2i) = 244.043 \ 648,$$

$$\Gamma(4+i)\Gamma(4-i) = 27.202 \ 292,$$

from which,  $K = 260.171 \ 676$ ,

$$f_4 = 260.171 \ 676 \frac{\Gamma(8+i)\Gamma(8-i)}{\Gamma(9+2i)\Gamma(9-2i)}.$$

Again making use of formula (13) we have

$$f_4 = 260.171 \ 676 \cdot \frac{22,243,314}{1,020,258,635} = 5.6722,$$

$$f_{15} = 260.171 \ 676 \frac{\Gamma(19+i)\Gamma(19-i)}{\Gamma(20+2i)\Gamma(20-2i)}.$$

Since  $R$  is fairly large in this instance, formula (17) is used and all terms of  $\psi(R, \varphi)$  after the first are dropped. This result gives

$$\log \Gamma(19+i)\Gamma(19-i) = 31.5892 \ 259,$$

$$\log \Gamma(20+2i)\Gamma(20-2i) = 34.0812 \ 782.$$



Accordingly,  $\log f_{15} = 9.9232\ 071 - 10$

and  $f_{15} = .8379$ .

By the same method  $f_{150}$  is calculated and we find  $f_{150} = .008723$ .

As a check on the accuracy of the results obtained in the above computations, values of  $f_x$  for  $x$  ranging from 1 to 15 were computed, using the given equation as a recursion formula. That is

$$f_1 = \frac{17}{29}f_0 = 17, \quad f_2 = \frac{26}{40}f_1 = 11.05, \quad \text{etc.}$$

These results are given in the following table, and it is to be noted that the values in the table for  $f_4$  and  $f_{15}$  agree with those previously computed by use of formulas contained in this paper. For obvious reasons, no attempt was made to compute the value of  $f_{150}$  by this method.

TABLE I

$x$	$f_x$	$x$	$f_x$	$x$	$f(x)$
0	29.0000	5	4.3375	10	1.6228
1	17.0000	6	3.4200	11	1.3961
2	11.0500	7	2.7633	12	1.2135
3	7.7142	8	2.2779	13	1.0644
4	5.6722	9	1.9092	14	0.9411
				15	0.8379

## REFERENCES

- [1] H. C. CARVER, "On the Graduation of Frequency Distributions," *Proc. of the Casualty Actuarial Soc. of America*, Vol. 6 (1919), pp. 52-72.
- [2] GULDBERG AND WALLENBERG, *Differenzengleichungen*, B. G. Teubner 1911, p. 25.
- [3] WHITTAKER AND WATSON, *Modern Analysis*, Cambridge Univ. Press 2nd Ed. 1911, pp. 229-258.
- [4] LIPSCHITZ, "Ueber die Darstellung gewisser Functionen durch die Eulersche Summenformel," *Crelle's Journal*, Bd. 56, s. 20.
- [5] "Preliminary Report on Computation of  $G_{(r)}$  Integrals," *Reports of British Assn. for Adv. Science*, 1896, p. 70.

UNIVERSITY OF MICHIGAN,  
ANN ARBOR, MICH.

**COMPARISON OF PEARSONIAN APPROXIMATIONS WITH EXACT  
SAMPLING DISTRIBUTIONS OF MEANS AND VARIANCES  
IN SAMPLES FROM POPULATIONS COMPOSED OF  
THE SUMS OF NORMAL POPULATIONS**

BY G. A. BAKER.

**1. Introduction.** Biological and sociological data are often "non-homogeneous" and of such a nature as not to be easily separated into components. Non-homogeneous populations have been discussed by Karl Pearson, Charlier, and others. Non-normal material has been discussed by many writers. See for example, A. E. R. Church [1] and J. M. LeRoux [2] for a discussion of moments of the distributions of the means and variances for samples from non-normal material.

In a previous paper [3] the author has given the distributions of the means and standard deviations of samples from certain non-homogeneous populations. The purpose of the present paper is to extend the results given in [3] and to compare the moment approach of the Pearsonian school with the true distributions.

**2. Moments of the distribution of means of samples of  $n$  from a non-homogeneous population.** Consider a population with distribution

$$(2.1) \quad f(x) = \frac{1}{(1+k)\sqrt{2\pi}} \left[ e^{-\frac{1}{2}\sigma^2 x^2} + \frac{k}{\sigma} e^{-\frac{1}{2\sigma^2}(x-m)^2} \right].$$

The first four moments of (2.1) about  $x = 0$  are

$$(2.2) \quad \begin{aligned} v'_1 &= \frac{km}{1+k} \\ v'_2 &= \frac{1}{1+k} [1 + k(\sigma^2 + m^2)] \\ v'_3 &= \frac{km}{1+k} [3\sigma^2 + m^2] \\ v'_4 &= \frac{1}{1+k} [3 + k(3\sigma^4 + 6m^2\sigma^2 + m^4)]. \end{aligned}$$

The means of samples of  $n$  drawn at random from (2.1) are distributed according to

$$(2.3) \quad \frac{n}{\sqrt{2\pi}(1+k)^n} \sum_{s=0}^n \binom{n}{s} \frac{k^s}{\sqrt{s\sigma^2 + n - s}} \exp \left\{ -\frac{n^2 \left( x - \frac{s}{n} m \right)^2}{s\sigma^2 + n - s} \right\}$$

Denote by  $m'_p$  the moments of (2.3) about  $x = 0$  and by  $m_p$  the moments about the mean. Then in view of the relations

$$(2.4) \quad \frac{n! s^r}{(n-s)! s!} = \sum_{i=0}^{r-1} A_{ri} \frac{n!}{(n-s)! (s-r+i)!}$$

$$A_{r0} = 1, \quad A_{r(r-1)} = 1, \quad A_{31} = 3, \quad A_{41} = 6, \quad A_{42} = 7,$$

$$A_{51} = 10, \quad A_{52} = 25, \quad A_{53} = 15,$$

and similar relations, and reduction to moments about the mean we obtain

$$(2.5) \quad m'_1 = \frac{km}{1+k} = v'_1$$

$$m_2 = \frac{1}{n(1+k)} \left[ 1 + k\sigma^2 + \frac{km^2}{1+k} \right]$$

$$m_3 = \frac{km}{n^2(1+k)^2} \left[ 3\sigma^2 - 3 + \frac{1-k}{1+k} m^2 \right]$$

$$m_4 = \frac{1}{n^3(1+k)^2} \left[ 3k(nk+1)\sigma^4 + 3(n+k) + 6(n-1)k\sigma^2 \right. \\ \left. + \frac{6k}{1+k} \{k + (n-1)\} m^2 + \frac{6k}{1+k} \{(n-1)k+1\} m^2 \sigma^2 \right. \\ \left. + \frac{k}{(1+k)^2} \{k^2 + (3n-4)k+1\} m^4 \right]$$

$$m_5 = \frac{k}{n^4(1+k)^3} \left[ 15\{(2n-1)k+1\} m\sigma^4 - 15\{k + (2n-1)\} m \right. \\ \left. + 30(n-1)(1-k)m\sigma^2 \right. \\ \left. + \frac{10}{1+k} \{-(n-1)k^2 + 4(n-1)k+1\} m^3 \sigma^2 \right. \\ \left. + \frac{10}{1+k} \{-k^2 - 4(n-1)k + (n-1)\} m^3 \right. \\ \left. + \frac{1}{(1+k)^2} \{-k^3 + (-10n+11)k^2 + (10n-11)k+1\} m^5 \right].$$

The expressions for the first five moments agree with the results given by Church and Tchebycheff.

The betas of (2.4) are

$$(2.6) \quad {}_1B_1 = \frac{k^2 m^2}{n(1+k)} \left[ \frac{3\sigma^2 - 3 + \frac{1-k}{1+k} m^2}{k\sigma^2 + 1 + \frac{k}{1+k} m^2} \right]^2$$

$$(2.7) \quad {}_1B_2 - 3 = \frac{k \left[ 3 + 3\sigma^4 - 6\sigma^2 - \frac{6}{1+k} m^2 + \frac{6}{1+k} m^2 \sigma^2 + \frac{k^2 - 4k + 1}{(1+k)^2} m^4 \right]}{n \left[ k\sigma^2 + 1 + \frac{k}{1+k} m^2 \right]^2}.$$

${}_1B_1$  vanishes if  $k = 0$ ,  $m = 0$ , or  $k = 1$  and  $\sigma = 1$ . If  $k$  and  $\sigma$  are constant and  $m$  approaches infinity  ${}_1B_1$  approaches  $(1 - k)^2/nk$ . If  $k$  and  $m$  are constant and  $\sigma$  approaches infinity  ${}_1B_1$  approaches zero.  ${}_1B_2 - 3$  vanishes if  $k = 0$ ,  $k = \infty$ , or if  $m = 0$  and  $\sigma = 1$ . If  $k$  and  $\sigma$  are constant and  $m$  approaches

TABLE I

$m_3$  and  ${}_pm_3$  compared for four sets of values of  $k$ ,  $\sigma^2$ , and  $m$

Sets of values $k \quad \sigma^2 \quad m$	$m_3$	${}_pm_3$
1/2   1/4   1.1	$-\frac{4.599}{n^3} + \frac{1.228}{n^4}$	$\frac{1}{n^3} \left( -4.250 + \frac{.235}{n} - \frac{.115}{n^2} \right) \left( 1 + \frac{.148}{n} \right)$
1/3   1   3.2	$\frac{89.702}{n^3} - \frac{39.322}{n^4}$	$\frac{1}{n^3} \left( 71.313 - \frac{26.709}{n} - \frac{4.886}{n^2} \right) \left( 1 + \frac{.584}{n} \right)$
-1/4   1/4   0.5	$\frac{4.640}{n^3} - \frac{1.744}{n^4}$	$\frac{1}{n^3} \left( 4.746 - \frac{.095}{n} - \frac{.165}{n^2} \right) \left( 1 + \frac{.199}{n} \right)$
1   4   5.6	$\frac{1,302.840}{n^3} - \frac{646.060}{n^4}$	$\frac{1}{n^3} \left( 762.035 - \frac{347.204}{n} - \frac{2.405}{n^2} \right) \left( 1 + \frac{.277}{n} \right)$

infinity then  ${}_1B_2 - 3$  approaches  $(k^2 - 4k + 1)/nk$ . If  $k$  and  $m$  are constant and  $\sigma$  approaches infinity then  ${}_1B_2 - 3$  approaches  $3/nk$ .

It is of interest to compare the higher moments of (2.3) with the higher moments calculated from the first four moments on the assumption of a Pearson curve in place of (2.3). On this assumption

$$(2.8) \quad {}_pm_3 = \frac{2m_3(m_4 + 7m_2^2m_4 - 3m_2m_3^2)}{9m_2^3 - m_2m_4 + 3m_3^2}.$$

It is seen that (2.8) bears little resemblance to  $m_3$ . If we consider the difference  ${}_pm_3 - m_3$  we see that it is of the same order in  $1/n$  as is  $m_3$  and the

numerator is of the 16th degree in  $k$ ,  $m$ , and  $\sigma$ ; a very complicated locus.  $m_s$  and  ${}_pm_s$  are compared for certain values of the parameters of (2.1) in Table I.

Table I shows that the coefficients of  $1/n^3$  in the expressions for  $m_s$  and  ${}_pm_s$  differ by from two to more than 40 per cent. The coefficients of  $1/n^4$  differ even more. The assumption of Karl Pearson's curves to represent the distribution of means of samples of  $n$  from non-homogeneous populations seems to be adequate in some cases but inadequate in others even for moderate values of the parameters.

**3. Moments of the distribution of variances.** In [3] an estimate of  $n$  times the standard deviation squared is expressed as

$$(3.1) \quad \bar{W} = (n - s) \bar{\sigma}_1^2 + s \bar{\sigma}_2^2 + \frac{(n - s)s}{n} (\bar{m}_1 + \bar{m}_2)^2,$$

where a bar over a letter means an estimate of the corresponding population parameter and where  $(n - s)$  denotes the number drawn from the first component of (2.1) and  $s$  denotes the number from the second component.

For the direct calculation of the moments of the distribution of variances it is easier not to use the distribution given in [3], but to proceed as follows. Put

$$(n - s) \bar{\sigma}_1^2 = y, \quad s \bar{\sigma}_2^2 = x, \quad \frac{(n - s)s}{n} (\bar{m}_1 + \bar{m}_2)^2 = z.$$

Of course, for population (2.1)  $\sigma_1 = 1$ ,  $\sigma_2 = \sigma$ ,  $m_1 = 0$ ,  $m_2 = m$ . The variables,  $x$ ,  $y$ ,  $z$  are all independent in the probability sense and their probability distributions are well known. Hence the moments of

$$(3.2) \quad \frac{\bar{W}}{n} = \frac{x + y + z}{n}$$

can be directly calculated.

For instance, if  $p = 1$  then

$$(3.3) \quad M'_1 = \frac{n - 1}{n} \frac{1}{1 + k} \left[ k\sigma^2 + 1 + \frac{k}{1 + k} m^2 \right].$$

In general, of course, the moments about the mean check with the values given by Church.

It is generally recommended to represent the distributions of variances of samples from non-normal parents by Pearson's curves. Let us examine the results of this procedure in a special case.

Suppose that the sampled population is

$$(3.4) \quad f(x) = \frac{1}{2\sqrt{2\pi}} [e^{-\frac{1}{2}x^2} + e^{-\frac{1}{2}(x-3.4)^2}].$$

The first eight moments of (3.4) which are needed in the calculation of the first four moments of the variances are:

$$\begin{aligned}
 (3.5) \quad & v'_1 = 1.7000 & v_5 = 0 \\
 & v_2 = 3.8900 & v_6 = 294.47 \\
 & v_3 = 0 & v_7 = 0 \\
 & v_4 = 28.692 & v_8 = 3,818.4.
 \end{aligned}$$

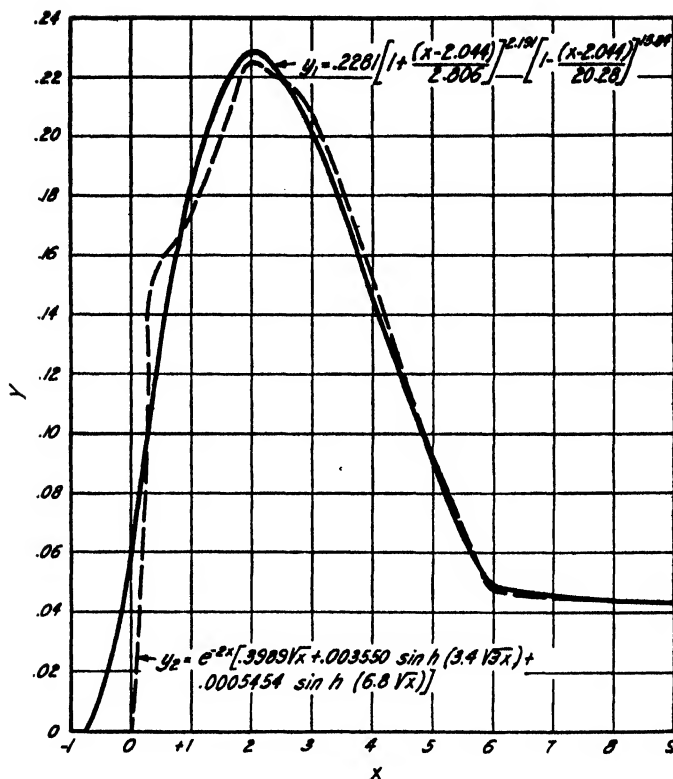


FIG. 1. COMPARISON OF THE TRUE DISTRIBUTION OF THE VARIANCES OF SAMPLES OF 4 DRAWN FROM THE NON-HOMOGENEOUS POPULATION (3.4) WITH THE CORRESPONDING EMPIRICAL PEARSON CURVE

The first four moments of the variances of samples of 4 from (3.4) are:

$$\begin{aligned}
 (3.6) \quad & {}_2M'_1 = 2.918 & {}_2M_3 = 4.745 \\
 & {}_2M_2 = 3.396 & {}_2M_4 = 41.52.
 \end{aligned}$$

Hence  ${}_2B_1 = .60$  and  ${}_2B_2 = 3.6$ ,  $\kappa = -.87$  which calls for a type 1 curve. The equation of the curve is

$$(3.7) \quad y_1 = 0.2281 \left(1 + \frac{x}{2.806}\right)^{2.191} \left(1 - \frac{x}{20.28}\right)^{15.84}$$

with its origin at its mode. The corresponding true distribution with the origin at the beginning of the range is

$$(3.8) \quad y_2 = e^{-2x} [.3989\sqrt{x} + .003550 \sinh (3.4\sqrt{3x}) \\ + .0005454 \sinh (6.8\sqrt{x})].$$

Distribution (3.8) differs slightly from the corresponding result given in [3] because of an error in that paper.

The two distributions are compared in Figure 1. It is seen that the two distributions are quite different. As the number of components of distributions similar to (3.8) increases, which is true as  $n$  increases, the distributions may be expected to become smoother and more closely representable by a single smooth curve.

**4. Summary.** The moments of the distribution of the means of samples of  $n$  from a non-homogeneous population composed of two normal components are given up to and including the fifth. This fifth moment is compared with the fifth moment calculated on the assumption of Pearson's curves to represent the distribution of means. The B's of the distributions of the means are discussed in certain limiting cases. It appears that for small samples and extreme values of the parameters, and in some cases of moderate values of the parameters, the Pearsonian approximations give poor results.

Some identities involving the binomial coefficients are given which permit the reduction of the moments of the distribution of means calculated directly to forms given elsewhere [1]. A method is given for the direct calculation of the moments of the variances of samples from a non-homogeneous population composed of two normal components. An indication of the closeness with which a Pearson curve can be made to fit the distribution of variances in small samples from a non-homogeneous population is given in Figure 1.

#### REFERENCES

- [1] A. E. R. CHURCH, "On the means and squared standard-deviations of small samples from any population," *Biometrika*, Vol. 18 (1926), pp. 321-394.
- [2] J. M. LEROUX, "A study of the distribution of the variance in small samples," *Biometrika*, Vol. 23 (1931), pp. 134-190.
- [3] G. A. BAKER, "Random Sampling from non-homogeneous populations," *Metron*, Vol. 8, no. 3, (1930).

UNIVERSITY OF CALIFORNIA,  
DAVIS, CALIF.

## A LEAST SQUARES ACCUMULATION THEOREM

BY W. E. BLEICK

The following simple least squares theorem does not seem to have been mentioned in the literature, and has at least one practical application.

If  $A^*(x)$  and  $B^*(x)$  are polynomials of the same degree which are least squares representations of the functions  $A(x)$  and  $B(x)$  respectively, for the values  $x_1, x_2, x_3, \dots, x_p$ , then

$$(1) \quad \sum_{i=1}^p A^*(x_i)B(x_i) = \sum_{i=1}^p A(x_i)B^*(x_i) = \sum_{i=1}^p A^*(x_i)B^*(x_i).$$

To prove the theorem let

$$(2) \quad A^*(x) = \sum_{i=0}^m a_i x^i$$

and

$$(3) \quad B^*(x) = \sum_{j=0}^n b_j x^j.$$

Then the normal equations for the determination of  $a_i$  and  $b_j$  are

$$(4) \quad \sum_{i=0}^m a_i s_{i+k} = \sum_{i=1}^p x_i^k A(x_i), \quad k = 0, 1, 2, \dots, m,$$

and

$$(5) \quad \sum_{j=0}^n b_j s_{j+h} = \sum_{i=1}^p x_i^h B(x_i), \quad h = 0, 1, 2, \dots, n,$$

where  $s_r = \sum_{i=1}^p x_i^r$ . Hence, by (2) and (5)

$$\begin{aligned} \sum_{i=1}^p A^*(x_i)B(x_i) &= \sum_{i=1}^p \left[ \sum_{i=0}^m a_i x_i^i \right] B(x_i) \\ &= \sum_{i=0}^m a_i \sum_{i=1}^p x_i^i B(x_i) \\ (6) \quad &= \sum_{i=0}^m \sum_{j=0}^n a_i b_j s_{i+j} \quad \text{if } n \geq m, \\ &= \sum_{i=1}^p A^*(x_i)B^*(x_i) \quad \text{if } n \geq m. \end{aligned}$$

Similarly it can be shown that

$$(7) \quad \sum_{i=1}^p A(x_i)B^*(x_i) = \sum_{i=1}^p A^*(x_i)B^*(x_i) \quad \text{if } m \geq n.$$



Combining (6) and (7) we have

$$(8) \quad \sum_{i=1}^p A^*(x_i)B(x_i) = \sum_{i=1}^p A(x_i)B^*(x_i) = \sum_{i=1}^p A^*(x_i)B^*(x_i) \quad \text{if } m = n.$$

In the particular case  $A(x) = B(x)$ , equation (8) gives the interesting result

$$(9) \quad \sum_{i=1}^p A^*(x_i)[A(x_i) - A^*(x_i)] = 0.$$

An obvious extension of equation (6) is

$$(10) \quad \sum_{i=1}^p x_i^q A^*(x_i)B(x_i) = \sum_{i=1}^p x_i^q A^*(x_i)B^*(x_i), \quad \text{if } n \geq m + q,$$

where  $q$  is a positive integer.

A practical application of (8) has been made by one large insurance company in the case  $m = n = 1$ . Suppose that  $A(x)$  represents an annual payment made  $x$  years ago and is an approximately linear function, and that  $B(x)$  represents a compound interest function. Then, even if  $B(x)$  is not a linear function, we may write approximately

$$\begin{aligned} \sum_{x=1}^p A(x)B(x) &\cong \sum_{x=1}^p A(x)B^*(x) \\ &\cong \sum_{x=1}^p A(x)(b_0 + b_1x) \\ (11) \quad &\cong b_0 \sum_{x=1}^p A(x) + b_1 \sum_{x=1}^p xA(x) \\ &\cong b_0 \sum_{x=1}^p A(x) + b_1 \sum_{x=1}^p \sum_{y=x}^p A(y). \end{aligned}$$

Thus if a year-by-year record is kept of the annual payments  $A(x)$ , the sum  $\sum_{x=1}^p A(x)$ , and the double sum  $\sum_{x=1}^p \sum_{y=x}^p A(y)$ , and if  $b_0$  and  $b_1$  are tabulated functions of  $p$ , equation (11) affords a convenient method of evaluating  $\sum_{x=1}^p A(x)B(x)$  approximately.

The author wishes to acknowledge that the case  $m = n = 1$  of equation (8) and the above application were brought to his attention by John K. Dyer.

## PARABOLIC TEST FOR LINKAGE

BY N. L. JOHNSON

**1. Introduction.** In this paper a problem in testing statistical hypotheses which has applications in genetics will be treated from the standpoint of the Neyman-Pearson approach. This approach has been developed in a series of papers, [4], [5], [6], [7], [8], [9], [10], to which the reader is referred for definitions of the concepts of a simple statistical hypothesis, critical regions, power function of a test with respect to alternative hypotheses, and that of a test unbiased in the limit employed in the present paper.

**2. Statement of Problem.** We shall consider  $M$  independent experiments, which will each yield results falling into one of the four categories described by the possible combinations of the 4 events  $a$ , not- $a$  (or  $\bar{a}$ ),  $b$ , and not- $b$  (or  $\bar{b}$ ) as set up in the following table.

	$a$	not- $a$	
$b$	$p_1$	$p_2$	$P_1$
not- $b$	$p_3$	$p_4$	$1 - P_1$
	$P_2$	$1 - P_2$	$1$

We shall assume that the marginal probabilities are known and have values  $P_1, 1 - P_1, P_2, 1 - P_2$  as shown in the table. Thus  $P_1$  = probability of event  $b$  happening whether event  $a$  occurs or not. It is obvious that if, further, the probability of a result falling in any one category or cell is fixed, then the other three cell probabilities will also be fixed. For if  $p_1, p_2, p_3, p_4$  be the four cell probabilities as shown in the table above, we must have

$$(1) \quad p_1 + p_2 = P_1; \quad p_1 + p_3 = P_2; \quad p_2 + p_4 = 1 - P_2.$$

Hence the values of the cell probabilities will be determined by a single parameter  $\theta$ , say, as follows

$$(2) \quad \begin{aligned} p_1 &= P_1 P_2 e^{\theta} & p_2 &= P_1 (1 - P_2 e^{\theta}) \\ p_3 &= P_2 (1 - P_1 e^{\theta}) & p_4 &= 1 - P_1 - P_2 + P_1 P_2 e^{\theta}. \end{aligned}$$

The range of values which  $\theta$  may take for the set of admissible hypotheses is found from the conditions

$$(3) \quad 0 \leq p_i \leq 1 \quad (i = 1, 2, 3, 4)$$

to be

$$(4) \quad -\infty < \theta \leq \min(-\log P_1, -\log P_2) \text{ if } P_1 + P_2 \leq 1$$

but

$$(5) \quad \log(P_1^{-1} + P_2^{-1} - P_1^{-1}P_2^{-1}) \leq \theta \leq \min(-\log P_1, -\log P_2) \text{ if } P_1 + P_2 \geq 1.$$

The hypothesis tested,  $H_0$ , is that  $\theta = 0$ , i.e. that the events  $a$  and  $b$  are independent. It will be noticed that  $H_0$  is a simple hypothesis, since it specifies the probability law of the observed variables completely. In fact, if  $m_i$  be the number of results out of our  $M$  experiments which are in the  $i$ th category, then  $m_1, m_2, m_3, m_4$  are our observed variables, and we have

$$(6) \quad P\{m_1 = m'_1, m_2 = m'_2, m_3 = m'_3, m_4 = m'_4 \mid H_0\} = \frac{M! p_{01}^{m'_1} p_{02}^{m'_2} p_{03}^{m'_3} p_{04}^{m'_4}}{m'_1! m'_2! m'_3! m'_4!}$$

where  $p_{0i}$  is the value of  $p_i$  when  $\theta = 0$ .

This is the conceptual model used in testing for linkage in two pairs of genes;  $H_0$  corresponds to the hypothesis "there is no linkage." Fuller explanations are given by Fisher [3]. It should be noted, however, that Fisher uses a parameter  $\theta$  corresponding to  $\frac{1}{2}e^\theta$  in this paper.

**3. Basis of Selection of Test.** The question now arises; what test shall we choose for the hypothesis  $H_0$ ? That is, what should the critical region  $w$  be to give us results as satisfactory as possible? The main aim must be to avoid errors, both of first and second kind, as far as possible. The first kind of error is subject to control, since the probability of the sample point  $E$  falling in  $w$  when  $H_0$  is true (which we shall denote by  $P\{E \in w \mid H_0\}$ ) can be determined approximately,  $H_0$  being simple. The critical region  $w$  is therefore chosen, if possible, to give a definite level of significance to the test associated with it. However, there will usually be many regions which will do this, and in order to decide which of them give more satisfactory results we consider  $(1 - P\{E \in w \mid H\})$ ; i.e. the probability of the second kind of error with respect to an alternative hypothesis  $H$ , the first kind of error being fixed.

In the present case  $H$  will be determined by  $\theta$  and so we may put  $P\{E \in w \mid H\} = \beta(w \mid \theta)$ , where  $\beta(w \mid \theta)$ , considered as a function of  $\theta$ , will be the power function of the test associated with the critical region  $w$ . We want  $w$  to be such that  $\beta(w \mid 0) = \alpha$ .  $\alpha$  being the fixed level of significance while  $\beta(w \mid \theta)$  is as large as possible.

It is also desirable that we should accept the hypothesis  $H_0$  more often when it is true than when any one of the alternative hypotheses ( $H$ ) is true. Ex-

pressed symbolically, this means that

$$(7) \quad \beta(w | 0) \leq \beta(w | \theta) \quad \text{for all } \theta \neq 0.$$

Any test satisfying the last condition is said to be *unbiased*.

If  $\beta$  and  $\frac{\partial \beta}{\partial \theta}$  are each continuous and differentiable functions of  $\theta$ , and we consider only those alternative hypotheses specified by suitably small values of  $\theta$ , sufficient conditions for the test to be unbiased will be

$$(8) \quad \left. \frac{\partial \beta}{\partial \theta} \right|_{\theta=0} = 0,$$

$$(9) \quad \left. \frac{\partial^2 \beta}{\partial \theta^2} \right|_{\theta=0} > 0.$$

According to the terminology recently adopted by Daly [1], the tests of which it is known only that they satisfy (8) and (9), are called *locally unbiased*.

If a region  $w$  could be found such that,  $v$  being any other region for which

$$(10) \quad \beta(w | 0) = \beta(v | 0), \quad \text{then} \quad \beta(w | \theta) \geq \beta(v | \theta)$$

for all  $\theta \neq 0$ , this would give a test which would be the best with respect to any alternative hypothesis. However, it has been shown by Neyman [4] that under certain conditions, which many probability laws satisfy, such a test will not exist. An attempt is therefore made to control the power of the test with respect to hypotheses specifying values of  $\theta$  near to 0; hoping that the powers of the tests so obtained with respect to the other hypotheses will behave in a satisfactory manner. Thus Neyman and Pearson [9] define an "unbiased test of Type A" as a test corresponding to a critical region  $w$  such that if  $v$  be any other region in the sample space  $W$  for which

$$(11) \quad \beta(w | 0) = \beta(v | 0) = \alpha$$

and

$$(12) \quad \left. \frac{\partial \beta(w | \theta)}{\partial \theta} \right|_{\theta=0} = \left. \frac{\partial \beta(v | \theta)}{\partial \theta} \right|_{\theta=0} = 0$$

then

$$(13) \quad \left. \frac{\partial^2 \beta(w | \theta)}{\partial \theta^2} \right|_{\theta=0} \geq \left. \frac{\partial^2 \beta(v | \theta)}{\partial \theta^2} \right|_{\theta=0}.$$

In the problem which I am treating the conditions

$$(14) \quad \beta(w | 0) = \alpha; \quad \left. \frac{\partial \beta(w | \theta)}{\partial \theta} \right|_{\theta=0} = 0$$

implied by (11) and (12) above cannot, in general, be satisfied, since the distribution is discontinuous, i.e.  $P\{E \in w | H_0\}$  is a discontinuous function of  $w$  and, in

fact, for a given sample size, has only a finite number of possible values, none of which need be equal to  $\alpha$ .

However, it may be possible to find a test of  $H_0$  of a type called "unbiased in the limit (as  $M$  increases)," based on the limiting form of the multinomial distribution which is a continuous function of  $w$ . The definition [6] of a test "unbiased in the limit" will be taken as follows:

Suppose we have a sequence  $(w_M)$  of critical regions,  $w_M$  corresponding to a sample of size  $M$ , such that

(i) for any  $M$ , if  $v_M$  be any region for which

$$(15) \quad \beta(w_M | 0) = \beta(v_M | 0)$$

and

$$(16) \quad \left. \frac{\partial \beta(w_M | \theta)}{\partial \theta} \right]_{\theta=0} = \left. \frac{\partial \beta(v_M | \theta)}{\partial \theta} \right]_{\theta=0}$$

then

$$(17) \quad \left. \frac{\partial^2 \beta(w_M | \theta)}{\partial \theta^2} \right]_{\theta=0} \geq \left. \frac{\partial^2 \beta(v_M | \theta)}{\partial \theta^2} \right]_{\theta=0}$$

(ii)

$$(18) \quad \lim_{M \rightarrow \infty} \beta(w_M | 0) = \alpha,$$

(iii) if

$$(19) \quad \vartheta = \sqrt{M} (\theta - 0) = \sqrt{M} \theta$$

$$(20) \quad \lim_{M \rightarrow \infty} \left. \frac{\partial \beta(w_M | \vartheta)}{\partial \vartheta} \right]_{\vartheta=0} = 0$$

then the test associated with this sequence of critical regions is unbiased in the limit. I shall call such a test a test of type  $A_\infty$ .

The reason for using  $\vartheta$  as the variable in condition (19) above is that, unless our sequence of critical regions has been very badly or unluckily chosen, we shall have

$$(21) \quad \lim_{M \rightarrow \infty} \beta(w_M | \theta) = 1 \quad (\theta \neq 0)$$

while, by (18),  $\lim_{M \rightarrow \infty} \beta(w_M | 0) = \alpha$  and so, in general,  $\lim_{M \rightarrow \infty} \frac{\partial \beta(w_M | \theta)}{\partial \theta}$  will not exist at  $\theta = 0$ . Hence we introduce  $\vartheta$ , termed the *normalized error*, and, keeping  $\vartheta$  constant (and hence making  $\theta$  tend to zero) we form  $\lim_{M \rightarrow \infty} \frac{\partial \beta(w_M | \vartheta)}{\partial \vartheta}$ .

In the next section will be obtained a test of  $H_0$  which is of type  $A_\infty$ .

**4. Derivation of Test.** The composition of a sample of  $M$  experiments is uniquely determined by the numbers of results  $m_1, m_2, m_3$  falling in the 1st,

2nd and 3rd categories respectively. Thus any sample may be represented by a point  $E(m)$  in a three-dimensional sample space  $W(m)$  with coordinate axes of  $m_1, m_2$ , and  $m_3$ . It will occasionally be convenient to represent the sample by a point in a three-dimensional space with other axes. The following sample spaces will be used.

$W(m)$ —space with coordinate axes of  $m_1, m_2, m_3$

$W(d)$ — “ “ “ “ “  $d_1, d_2, d_3$

$W(x)$ — “ “ “ “ “  $x_1, x_2, x_3$

$W(n)$ — “ “ “ “ “  $n_1, n_2, n_3$

where

$$(22) \quad d_i = m_i - Mp_{0i} \quad (i = 1, 2, 3, 4)$$

$$(23) \quad x_i = (m_i - Mp_{0i})/(Mp_{0i})^{\frac{1}{2}} \quad (i = 1, 2, 3, 4)$$

$$(24) \quad n_i = m_i/M \quad (i = 1, 2, 3, 4).$$

I shall use  $w_M$  indifferently to denote “the critical region corresponding to sample size  $M$ ” in any of the four sample spaces above;  $E$  indifferently to denote corresponding positions of the sample point in any of the four sample spaces: except in cases where confusion might arise, where I shall use  $w_M(m)$ ,  $w_M(d)$ ,  $w_M(x)$ ,  $w_M(n)$  and  $E(m)$ ,  $E(d)$ ,  $E(x)$ ,  $E(n)$ . When necessary the size of sample with which a point  $E$  is associated will be denoted by a subscript; e.g.  $E_M$ .

In finding a test of type  $A_\infty$  we shall need to consider the quantities  $\beta(w_M | 0)$ ,  $\left. \frac{\partial \beta(w_M | \vartheta)}{\partial \vartheta} \right|_{\vartheta=0}$ , and  $\left. \frac{\partial^2 \beta(w_M | \theta)}{\partial \theta^2} \right|_{\theta=0}$ , where  $\vartheta = \theta \sqrt{M}$ .

The probability law of the observed values  $m_1, m_2, m_3$  is discontinuous with respect to the points of the sample space  $W_M$ . For if  $E^0$  be a point which corresponds to integral values  $m_1^0, m_2^0, m_3^0$  of  $m_1, m_2, m_3$ ; subject to the restrictions

$$(25) \quad 0 \leq m_i^0 \quad (i = 1, 2, 3)$$

$$(26) \quad 0 \leq \sum_{i=1}^3 m_i^0 \leq M$$

then

$$(27) \quad P\{E_M = E^0 | \theta = 0\} = \frac{M! p_{01}^{m_1^0} p_{02}^{m_2^0} p_{03}^{m_3^0} p_{04}^{m_4^0}}{m_1^0! m_2^0! m_3^0! m_4^0!}$$

where

$$(28) \quad \sum_{i=1}^4 m_i^0 = M$$

and

$$(29) \quad \begin{aligned} p_{01} &= P_1 P_2 & p_{02} &= P_1(1 - P_2) \\ p_{03} &= P_2(1 - P_1) & p_{04} &= (1 - P_1)(1 - P_2) \end{aligned}$$

while if  $E^0$  be not such a point

$$(30) \quad P\{E_M = E^0 \mid \theta\} = 0$$

whatever the value of  $\theta$  may be. Now

$$(31) \quad \beta(w_M \mid \theta) = \sum_{w_M} \frac{M! p_1^{m_1} p_2^{m_2} p_3^{m_3} p_4^{m_4}}{m_1! m_2! m_3! m_4!}$$

where  $p_1, p_2, p_3, p_4$  are as defined in (2) above, and  $\sum$  denotes a finite summation over all points  $E'$  in  $w_M$  for which  $P\{E_M = E' \mid \theta\} \neq 0$ . Differentiating each side of (31) with respect to  $\theta$ , we get

$$(32) \quad \left. \frac{\partial \beta(w_M \mid \theta)}{\partial \theta} \right]_{\theta=0} = \sum_{w_M} \frac{M! p_{01}^{m_1} p_{02}^{m_2} p_{03}^{m_3} p_{04}^{m_4}}{m_1! m_2! m_3! m_4!} \times \left[ \frac{m_1(1 - P_1 - P_2) - m_2 P_2 - m_3 P_1 + m_4 P_1 P_2}{(1 - P_1)(1 - P_2)} \right]$$

and

$$(33) \quad \left. \frac{\partial^2 \beta(w_M \mid \theta)}{\partial \theta^2} \right]_{\theta=0} = \sum_{w_M} \frac{M! p_{01}^{m_1} p_{02}^{m_2} p_{03}^{m_3} p_{04}^{m_4}}{m_1! m_2! m_3! m_4!} \cdot \frac{1}{(1 - P_1)^2 (1 - P_2)^2} [\{m_1(1 - P_1 - P_2) - m_2 P_2 - m_3 P_1 + M P_1 P_2\}^2 - \{m_1 P_1 P_2(1 - P_1 - P_2) + m_2 P_2(1 - P_1 - P_1 P_2) + m_3 P_1(1 - P_2 - P_1 P_2) - M P_1 P_2(1 - P_1)(1 - P_2)\}].$$

THEOREM 1. The sequence of critical regions ( $w_M$ ) defined by

$$(34) \quad v + B u^2 \geq A \text{ in } w_M; \quad v + B u^2 < A \text{ elsewhere,}$$

where

$$(35) \quad u = \frac{x_1(P_1 P_2)^{\frac{1}{2}}(1 - P_1 - P_2) - x_2 P_1^{\frac{1}{2}}(1 - P_2)^{\frac{1}{2}} P_2 - x_3 P_2^{\frac{1}{2}}(1 - P_1)^{\frac{1}{2}} P_1}{\{P_1 P_2(1 - P_1)(1 - P_2)\}^{\frac{1}{2}}}$$

$$(36) \quad v = \frac{P_1(1 - P_1)(2P_2 - 1)\{x_1(P_1 P_2)^{\frac{1}{2}} + x_3 P_2^{\frac{1}{2}}(1 - P_1)^{\frac{1}{2}}\} + P_2(1 - P_2)(2P_1 - 1)\{x_1(P_1 P_2)^{\frac{1}{2}} + x_2 P_1^{\frac{1}{2}}(1 - P_2)^{\frac{1}{2}}\}}{[P_1 P_2(1 - P_1)(1 - P_2)\{P_1(1 - P_1)(1 - 2P_2)^2 + P_2(1 - P_2)(1 - 2P_1)^2\}]^{\frac{1}{2}}}$$

$$(37) \quad B = \left[ \frac{M P_1 P_2(1 - P_1)(1 - P_2)}{P_1(1 - P_1)(1 - 2P_2)^2 + P_2(1 - P_2)(1 - 2P_1)^2} \right]^{\frac{1}{2}}$$

$$(38) \quad \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}u^2} \int_{-\beta u^2}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du = \alpha$$



$$(43) \quad |\beta(w|0) - I(w)| < \epsilon$$

where

$$(44) \quad I(w) = \frac{1}{(2\pi)^{\frac{1}{2}} \sqrt{p_{04}}} \iiint_{w(x)} e^{-ix_1^2} dx_1 dx_2 dx_3$$

and

$$(45) \quad \chi_0^2 = \sum_{i=1}^3 x_i^2 (1 + p_{0i} p_{0i}^{-1}) + 2 \sum_{i < j \leq 3} x_i x_j (p_{0i} p_{0j})^{\frac{1}{2}} p_{04}^{-1}.$$

We will now apply a transformation to the coordinates  $m_1, m_2, m_3$  which will

(a) transform inequality (42) into a simpler form,

(b) transform  $I(w)$  into a form to which the tables of the Normal Probability Integral may easily be applied for purposes of calculation.

This transformation is

$$(46) \quad u = \frac{x_1(P_1 P_2)^{\frac{1}{2}}(1 - P_1 - P_2) - x_2 P_1^{\frac{1}{2}}(1 - P_2)^{\frac{1}{2}} P_2 - x_3 P_2^{\frac{1}{2}}(1 - P_1)^{\frac{1}{2}} P_1}{\{P_1 P_2(1 - P_1)(1 - P_2)\}^{\frac{1}{2}}}$$

$$(47) \quad v = \frac{P_1(1 - P_1)(2P_2 - 1)\{x_1(P_1 P_2)^{\frac{1}{2}} + x_3 P_2^{\frac{1}{2}}(1 - P_1)^{\frac{1}{2}}\} + P_2(1 - P_2)(2P_1 - 1)\{x_1(P_1 P_2)^{\frac{1}{2}} + x_2 P_1^{\frac{1}{2}}(1 - P_2)^{\frac{1}{2}}\}}{[P_1 P_2(1 - P_1)(1 - P_2)\{P_1(1 - P_1)(1 - 2P_2)^2 + P_2(1 - P_2)(1 - 2P_1)^2\}]^{\frac{1}{2}}}$$

$$(48) \quad t = \frac{(2P_1 - 1)\{x_1(P_1 P_2)^{\frac{1}{2}} + x_3 P_2^{\frac{1}{2}}(1 - P_1)^{\frac{1}{2}}\} - (2P_2 - 1)\{x_1(P_1 P_2)^{\frac{1}{2}} + x_2 P_1^{\frac{1}{2}}(1 - P_2)^{\frac{1}{2}}\}}{\{P_1(1 - P_1)(1 - 2P_2)^2 + P_2(1 - P_2)(1 - 2P_1)^2\}^{\frac{1}{2}}}.$$

This is a proper transformation, since under the conditions of the theorem  $0 < P_i < 1$  and  $P_1$  and  $P_2$  are not both  $\frac{1}{2}$ ; and the Jacobian

$$(49) \quad J = \frac{\partial(u, v, t)}{\partial(x_1, x_2, x_3)} = p_{04}^{-1}$$

is non-zero and of constant sign.

Also

$$(50) \quad \chi_0^2 = u^2 + v^2 + t^2.$$

Hence

$$(51) \quad I(w) = \frac{1}{(2\pi)^{\frac{1}{2}}} \iiint_{w(u,v,t)} e^{-\frac{1}{2}(u^2+v^2+t^2)} du dv dt.$$

The inequality (42) is transformed into an inequality of form  $B(u - a_s)^2 + v \geq A$  where  $B$  has the value stated above;  $a_s$  and  $A$  being at present arbitrary constants.

Therefore we may put  $a_s = 0$  and define  $A$  by the equation

$$(52) \quad \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}u^2} \int_{A-Bu^2}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du = \alpha$$

and conclude that the sequence of critical regions  $(w_M)$  defined by the inequalities

$$(53) \quad \begin{aligned} Bu^2 + v &\geq A \quad \text{in } w_M \\ Bu^2 + v &< A \quad \text{elsewhere} \end{aligned}$$

will satisfy conditions (i) for a test of type  $A_\infty$ .

From (51) and (52)

$$(54) \quad \begin{aligned} I(w_M) &= \frac{1}{(2\pi)^4} \iiint_{w_M} e^{-i(u^2+v^2+t^2)} du dv dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ e^{-iu^2} \int_{A-Bu^2}^{\infty} e^{-iv^2} dv \right\} du = \alpha. \end{aligned}$$

By THEOREM 1 of the appendix, as mentioned above, we have

$$(55) \quad |\beta(w_M | 0) - I(w_M)| < \epsilon \quad \text{for all } M > M_\epsilon$$

i.e.

$$(56) \quad |\beta(w_M | 0) - \alpha| < \epsilon \quad \text{for all } M > M_\epsilon$$

and so

$$(57) \quad \beta(w_M | 0) \rightarrow \alpha \quad \text{as } M \rightarrow \infty.$$

Thus the sequence of critical regions  $(w_M)$  satisfies the condition (ii) of the definition of a test of type  $A_\infty$ .

If  $w$  be any region defined by inequalities on  $u$  and  $v$  only (as are the regions  $w_M$ ) then, as a special case of THEOREM 1 of the Appendix, we have that for any  $\epsilon > 0$  there exists a number  $M_\epsilon$  such that for all  $M > M_\epsilon$ ,

$$(58) \quad P_M(w) - \frac{1}{2\pi} \iint_{w(u,v)} e^{-i(u^2+v^2)} du dv < \epsilon$$

where  $P_M(w) = P\{E_M \in w | 0\}$ .

By (31) and (32), noting that  $\frac{\partial \beta(w | \theta)}{\partial \theta} = \sqrt{M} \cdot \frac{\partial \beta(w | \vartheta)}{\partial \vartheta}$ , we have

$$(59) \quad \begin{aligned} \left. \frac{\partial \beta(w | \vartheta)}{\partial \vartheta} \right]_{\vartheta=0} &= \sum_w f_1(u, v) \cdot u \cdot (P_1 P_2)^{\frac{1}{2}} (1 - P_1)^{-1} (1 - P_2)^{-1} \\ &= \sum_w f_1(u, v) \cdot uk \end{aligned}$$

where  $k = (P_1 P_2)^{\frac{1}{2}} (1 - P_1)^{-1} (1 - P_2)^{-1} > 0$ .

By THEOREM 1 of the Appendix, as last stated above, we have

$$(60) \quad f_1(u, v) = \frac{1}{2\pi} \Delta u \Delta v \cdot e^{-i(u^2+v^2)} (1 + R_M)$$

where for convenience we have written  $\Delta u$ ,  $\Delta v$  for  $\Delta_{(M)} u$ ,  $\Delta_{(M)} v$  the units of  $u$  and  $v$  when sample size is  $M$ , and  $R_M$  for  $R_M(u, v)$  which has the property that

$$(61) \quad \sum_w R_M(u, v) \Delta_{(M)} u \cdot \Delta_{(M)} v \cdot e^{-\frac{1}{2}(u^2+v^2)} \rightarrow 0$$

uniformly with respect to  $w$  as  $M \rightarrow \infty$ .

Now let  $w^+$  denote that part of  $w$  where  $R_M \geq 0$  and  $w^-$  that part of  $w$  where  $R_M < 0$ . Then

$$(62) \quad \sum_{w^+} k u f_1(u, v) = \sum_{w^+} k \cdot \frac{\Delta u \Delta v}{2\pi} \cdot u e^{-\frac{1}{2}(u^2+v^2)} + \sum_{w^+} k \frac{\Delta u \Delta v}{2\pi} u R_M e^{-\frac{1}{2}(u^2+v^2)}.$$

Let

$$(63) \quad \begin{aligned} S_M^+ &= \sum_{w^+} k \frac{\Delta u \Delta v}{2\pi} u R_M e^{-\frac{1}{2}(u^2+v^2)} \\ &= k \sum_{w^+} \left\{ \left( R_M \frac{\Delta u \Delta v}{2\pi} \right)^{\frac{1}{2}} u e^{-\frac{1}{4}(u^2+v^2)} \right\} \left\{ \left( R_M \frac{\Delta u \Delta v}{2\pi} \right)^{\frac{1}{2}} e^{-\frac{1}{4}(u^2+v^2)} \right\}. \end{aligned}$$

By Schwarz's inequality

$$(64) \quad \frac{S_M^+}{k} \leq \sum_{w^+} \frac{\Delta u \Delta v}{2\pi} u^2 R_M e^{-\frac{1}{2}(u^2+v^2)} \leq \sum_{w^+} \frac{\Delta u \Delta v}{2\pi} R_M e^{-\frac{1}{2}(u^2+v^2)}$$

But

$$(65) \quad \sum_{w^+} u^2 f_1(u, v) = \sum_{w^+} \frac{\Delta u \Delta v}{2\pi} u^2 e^{-\frac{1}{2}(u^2+v^2)} + \sum_{w^+} \frac{\Delta u \Delta v}{2\pi} u^2 R_M e^{-\frac{1}{2}(u^2+v^2)}.$$

Now  $u^2 f_1(u, v) \geq 0$  and  $\sum_w u^2 f_1(u, v)$  is finite (since  $u^2$  is a homogeneous function of second degree in the  $x_i$ 's and so has a finite expectation) and is bounded as  $M \rightarrow \infty$ . Hence  $\sum_{w^+} u^2 f_1(u, v)$  is finite and bounded as  $M \rightarrow \infty$ . Further, as  $M \rightarrow \infty$

$$(66) \quad \sum_{w^+} \frac{\Delta u \Delta v}{2\pi} u^2 e^{-\frac{1}{2}(u^2+v^2)} \rightarrow \frac{1}{2\pi} \iint u^2 e^{-\frac{1}{2}(u^2+v^2)} du dv.$$

Hence  $\sum_{w^+} \frac{\Delta u \Delta v}{2\pi} u^2 R_M e^{-\frac{1}{2}(u^2+v^2)}$  is bounded as  $M \rightarrow \infty$ . From this result, together with (61) and (64) it follows that  $S_M^+ \rightarrow 0$  as  $M \rightarrow \infty$  uniformly with respect to  $w$ . Putting

$$(67) \quad S_M^- = \sum_{w^-} k \frac{\Delta u \Delta v}{2\pi} u R_M e^{-\frac{1}{2}(u^2+v^2)}$$

it will follow in a similar manner that  $S_M^- \rightarrow 0$  as  $M \rightarrow \infty$  uniformly with respect to  $w$ . Hence

$$(68) \quad \begin{aligned} \left[ \frac{\partial \beta(w | \vartheta)}{\partial \vartheta} \right]_{\vartheta=0} &= \sum_w k u f_1(u, v) \\ &= \sum_w k \frac{\Delta u \Delta v}{2\pi} u e^{-\frac{1}{2}(u^2+v^2)} + S_M \end{aligned}$$

where  $S_M = S_M^+ + S_M^-$  and so  $S_M \rightarrow 0$  as  $M \rightarrow \infty$  uniformly with respect to  $w$ .

Hence whatever be  $\epsilon > 0$ , there is a number  $M'$ , such that for all  $M > M'$ ,

$$(69) \quad \left| \frac{\partial \beta(w|\vartheta)}{\partial \vartheta} \right]_{\vartheta=0} - \frac{k}{2\pi} \iint_w u e^{-i(u^2+v^2)} du dv \right| < \epsilon$$

whatever be the region  $w$ . In particular we may take  $w = w_M$ , and then we have

$$(70) \quad \frac{k}{2\pi} \iint_{w_M} u e^{-i(u^2+v^2)} du dv = \frac{k}{2\pi} \int_{-\infty}^{+\infty} \left\{ u e^{-iu^2} \int_{A-Bu^2}^{\infty} e^{-iv^2} dv \right\} du = 0$$

and so

$$(71) \quad \left| \frac{\partial \beta(w_M|\vartheta)}{\partial \vartheta} \right]_{\vartheta=0} \right| < \epsilon \quad \text{for all } M > M'$$

i.e.,

$$(72) \quad \lim_{M \rightarrow \infty} \frac{\partial \beta(w_M|\vartheta)}{\partial \vartheta} \Big|_{\vartheta=0} = 0.$$

Hence the sequence of critical regions ( $w_M$ ) satisfies condition (iii) for a test of type  $A_\infty$ . This completes the proof of THEOREM 1.

In the above theorem we have found a test which is unbiased in the limit for all cases except that for which  $P_1 = P_2 = \frac{1}{2}$ . The following theorem derives the test appropriate to this special case, and it is found that in this instance the test takes a very simple form.

**THEOREM 2.** *If  $P_1 = P_2 = \frac{1}{2}$ , the sequence of critical regions ( $w_M$ ) defined by*

$$(73) \quad \begin{aligned} |x_2 + x_3| &\geq a && \text{in } w_M \\ |x_2 + x_3| &< a && \text{elsewhere} \end{aligned}$$

where

$$(74) \quad \frac{1}{\sqrt{2\pi}} \int_a^{+\infty} e^{-\frac{1}{2}x^2} dx = 1 - \alpha$$

$$(75) \quad x_i = \frac{m_i - \frac{1}{2}M}{\frac{1}{2}M^{\frac{1}{2}}} \quad (i = 2, 3),$$

is associated with a test of the hypothesis  $H_0(\theta = 0)$  of type  $A_\infty$  at level of significance  $\alpha$ .

The proof of this theorem follows the same lines as that of Theorem 1 as far as inequality (42). On putting  $P_1 = P_2 = \frac{1}{2}$  in (42) we get

$$(76) \quad (-\frac{1}{2}m_2 - \frac{1}{2}m_3 + \frac{1}{2}M - a_0)^2 - \frac{1}{4}(m_2 + m_3 - \frac{1}{2}M) \geq a_1$$

i.e.,

$$(77) \quad (x_2 + x_3 - a_0)^2 \geq a_1.$$

The critical region  $w_M$  defined in the statement of the theorem is of this form with  $a_6 = 0$  and  $a_7 = a^{\frac{1}{2}}$ .

Hence the sequence of critical regions ( $w_M$ ) satisfies conditions (i) of the definition of a test of type  $A_\infty$ . The sequence of critical regions may also be shown to satisfy conditions (ii) and (iii) for a test of type  $A_\infty$  by following the lines of the proof of THEOREM 1 and noting that  $x_2 + x_3 = 2M^{-\frac{1}{2}}(m_2 + m_3 - \frac{1}{2}M)$  tends to be distributed as a unit normal deviate as  $M \rightarrow \infty$ .

On account of the shape of the critical regions in the general case, I shall for the remainder of this paper call the tests derived in the above theorem the *parabolic tests* for the cases considered.

**5. Application of the Parabolic Tests.** For practical purposes the formulae derived above are inconvenient to use. I will therefore express them in terms of the deviations of the observed frequencies in the four cells from the frequencies "expected" when the hypothesis  $H_0(\theta = 0)$  is true, i.e. in terms of the variables  $d_i$ , where

$$(78) \quad d_i = m_i - Mp_{0i} = x_i(Mp_{0i})^{\frac{1}{2}} \quad (i = 1, 2, 3, 4).$$

The test then becomes "reject the hypothesis  $H_0$  at level of significance  $\alpha$  if  $v + Bu^2 \geq A$ " where

$$(79) \quad u = \frac{d_1(1 - P_1 - P_2) - d_2P_2 - d_3P_1}{\{MP_1P_2(1 - P_1)(1 - P_2)\}^{\frac{1}{2}}}$$

$$(80) \quad v = \frac{P_1(1 - P_1)(2P_2 - 1)(d_1 + d_3) + P_2(1 - P_2)(2P_1 - 1)(d_1 + d_2)}{[MP_1P_2(1 - P_1)(1 - P_2)\{P_1(1 - P_1)(2P_2 - 1)^2 + P_2(1 - P_2)(2P_1 - 1)^2\}]^{\frac{1}{2}}}$$

$$(81) \quad \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}u^2} \int_{A-Bu^2}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du = \alpha$$

$$(82) \quad B = \left[ \frac{MP_1P_2(1 - P_1)(1 - P_2)}{P_1(1 - P_1)(1 - 2P_2)^2 + P_2(1 - P_2)(1 - 2P_1)^2} \right]^{\frac{1}{2}}$$

except when  $P_1 = P_2 = \frac{1}{2}$ . In the latter case reject the hypothesis  $H_0$  if

$$(83) \quad d_2 + d_3 \geq a$$

where

$$(84) \quad \frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} e^{-\frac{1}{2}x^2} dx = 1 - \alpha.$$

The application of this last case ( $P_1 = P_2 = \frac{1}{2}$ ) is straightforward.  $a$  may be found from the tables of the Normal Probability Integral.  $d_2$  and  $d_3$  may be

calculated from the data, and we may then see whether the inequality (83) is satisfied, and so assess our judgment of the hypothesis  $H_0$ .

TABLE I

*Significance of Symbols*

$A$  and  $B$  are connected by the following relation:

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}u^2} \int_{A-Bu^2}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du = \alpha.$$

Table Ia $\alpha = 0.05$ $p_{.05} = A - 3.8414588 B$		Table Ib $\alpha = 0.01$ $p_{.01} = A - 6.6348966 B$	
$B$	$p_{.05}$	$B$	$p_{.01}$
0	1.6449	0	2.3263
1.00	0.322	1.00	0.289
1.25	.256	1.25	.231
1.50	.212	1.50	.192
1.75	.181	1.75	.165
2.00	.158	2.00	.144
2.25	.141	2.25	.128
2.50	.127	2.50	.115
2.75	.116	2.75	.105
3.00	.106	3.00	.096
3.25	.098	3.25	.089
3.50	.091	3.50	.082
3.75	.084	3.75	.077
4.00	.079	4.00	.072
5	.063	5	.058
6	.052	6	.048
7	.045	7	.041
8	.039	8	.036
9	.035	9	.032
10	.031	10	.029
15	.021	15	.020
20	.016	20	.014
30	.010	30	.009
40	.008	40	.007
50	.006	50	.006

The general case is also straightforward, except for the determination of  $A$  from equation (81). To facilitate this I have constructed Tables Ia and Ib. These tables correspond respectively to significance levels .05, .01, and from

them the value of  $A$  corresponding to a given value of  $B$  may be calculated. The quantity tabled,  $(\rho)$ , is the difference between  $A$  and a multiple<sup>1</sup> (constant for a given level of significance and given with the table to which it applies) of  $B$ . To find  $A$ , therefore,  $B$  is calculated, multiplied by the appropriate constant, and added to the quantity in the table corresponding to  $B$ . For large values of  $B$  (40 and over)  $\rho$  is small, and  $A$  may be taken equal to the constant multiple of  $B$ .

In particular cases when the values of  $P_1$  and  $P_2$  are substituted in the expression for  $B$  (see THEOREM 1 above) and in (79) and (80) above, these equations appear much less formidable. Thus in the case considered by R. A. Fisher [3],  $P_1 = P_2 = \frac{1}{2}$  and we get

$$(85) \quad B = \sqrt{\frac{3M}{8}}$$

$$u = \frac{1}{2}M^{-1}(2d_1 - d_2 - d_3); \quad v = -4(6M)^{-1}(2d_1 + d_2 + d_3)$$

and the test becomes "reject the hypothesis  $H_0$  at level of significance  $\alpha$  when

$$(86) \quad \phi = \{(2d_1 - d_2 - d_3)^2 - \frac{1}{3}(2d_1 + d_2 + d_3)\} / \{\frac{1}{3}(\frac{1}{2}M)^1\} \geq A$$

where

$$(87) \quad \frac{1}{2\pi} \int_{-\infty}^{\infty} \left\{ e^{-\frac{1}{2}u^2} \int_{A-u^2\sqrt{1/M}}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du = \alpha.$$

*Example.* Fisher [3] gives an example of the case  $P_1 = P_2 = \frac{1}{2}$ . In the series of experiments that he quotes the observed results fall in the four categories respectively as follows:

$$m_1 = 32; \quad m_2 = 904; \quad m_3 = 906; \quad m_4 = 1997. \quad M = 3839.$$

Hence  $d_1 = -207.9375$ ;  $d_2 + d_3 = 370.375$ . From (86),  $\phi = 10863.1$ .  $B = 37.94239$ . From the tables:

$$\text{at .05 level, } A_{.05} = 3.8414588 \times 37.94239 + 0.0075 = \underline{145.7615}$$

$$\text{at .01 level, } A_{.01} = 6.6348966 \times 37.94239 + 0.0065 = \underline{251.750}.$$

Hence we reject the hypothesis that  $\theta = 0$ , i.e. that there is no linkage, since the value of  $\phi$  is well outside even the .01 level of significance.

**6. Power function of the Tests. General Case.** The parabolic test as described above has the desirable property that of all tests (at level of significance  $\alpha$ ) which are unbiased for large values of  $M$  this test will detect small variations in  $\theta$  most frequently. However, to get a clearer idea of the properties of this

<sup>1</sup> This multiple is equal to  $k\frac{1}{2}$  where  $\frac{1}{\sqrt{2\pi}} \int_{-k}^{+k} e^{-t^2} dt = 1 - \alpha$ ,  $\alpha$  being the level of significance.

test we shall calculate, as accurately as may be practicable, the power function of the test.

As a preliminary step we obtain a rough idea of the power function by making use of the concept of a limiting power function as stated by Neyman [6]. This may be defined as follows:

Let  $E_{M'}$  denote the sample point corresponding to a sample of size  $M'$ , and put

$$(88) \quad P\{E_{M'} \in w \mid \vartheta'\} = \beta_{M'}(w \mid \vartheta'),$$

where  $\vartheta' = M'^{1/2}\theta$ ,  $w$  being a fixed region. Supposing  $\vartheta'$  kept fixed, let  $M'$  increase and let

$$(89) \quad \beta_{\infty}(w \mid \vartheta') = \lim_{M' \rightarrow \infty} \beta_{M'}(w \mid \vartheta')$$

if this limit exists.

Then  $\beta_{\infty}(w \mid \vartheta')$  is the limiting power function of the test associated with the critical region  $w$ . It will be noted that the limiting power function is a function of  $\vartheta'$ .

In the problem under consideration the parabolic test when the sample size is  $M$  is associated with the critical region  $w_M$ . Now it should be noted that in the definition of the limiting power function  $w$  remains fixed. Therefore the limiting power function of the parabolic test for sample size  $M$  is

$$(90) \quad \beta^{\infty}(w_M \mid \vartheta') = \lim_{M' \rightarrow \infty} \beta_{M'}(w_M \mid \vartheta').$$

The significance of the limiting power function is that for any  $\epsilon > 0$  and for any  $\vartheta'$  there is a number  $M_{\epsilon, \vartheta'}$  such that for all  $M > M_{\epsilon, \vartheta'}$  we have in our case (by THEOREM 1 of the Appendix)

$$(91) \quad |\beta_M(w_M \mid \vartheta') - \beta_{\infty}(w_M \mid \vartheta')| < \epsilon.$$

It should be noted, however, that the limiting power curve (the graph of the limiting power function against  $\theta = \vartheta M^{-1/2}$ ) may be only a very rough approximation to the actual power curve. Furthermore (Neyman, [6, p. 83]) we cannot, in general, use the limiting power function of a test to answer the question:

"How large must we take our sample size  $M$  to detect the falsehood of the hypothesis  $H_0(\theta = 0)$  when actually  $\theta = \theta'$ , with a limiting probability of at least, say, 0.95?"

For if we form a table as below

$M$	$\vartheta'_{(M)} = M^{1/2}\theta'$	$\beta_{\infty}(w_M \mid \vartheta'_{(M)})$
100	...	...
1000	...	...

it is possible that  $\beta_{\infty}(w_M \mid \vartheta'_{(M)})$  may never attain the value 0.95.



THEOREM 3. *The limiting power function of the parabolic test is*

$$(92) \quad \beta_{\infty}(w_M | \theta) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}[u - \theta(P_1 P_2)^{\frac{1}{2}}(1-P_1)^{-\frac{1}{2}}(1-P_2)^{-\frac{1}{2}}]^2} \int_{A-Bu^2}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du$$

in all cases for which  $0 < P_i < 1$  and  $P_1$  and  $P_2$  are not both equal to  $\frac{1}{2}$ .

The proof of this theorem follows immediately from THEOREM 1 of the Appendix by applying the transformation (46)–(48) and putting  $\lambda = P_1 P_2$ .

The above remarks concerning special precautions to be taken with respect to the limiting power function suggest the necessity of studying the actual power function of the parabolic test by some other method.

With this object in view, a study was made of the distribution of the function  $\phi = v + Bu^2$  for finite values of  $M$  and in particular for  $M = 100$  and  $M = 3839$ .  $\phi$  is a discontinuous variate and, for any given value of  $M$ , has definite limits of variation arising from the limitations on the values of the variables  $m_i$  stated in the inequalities (25), (26) above. These limits of variation of  $\phi$  were found to be

$$(93) \quad -\frac{4}{3}(\frac{2}{3}M)^{\frac{1}{2}}(\frac{2}{3}M - \frac{1}{16}) < \phi < \frac{4}{3}(\frac{2}{3}M)^{\frac{1}{2}}M(\frac{2}{3}M - 1)$$

for the case  $P_1 = P_2 = \frac{1}{4}$ . Hence when

$$M = 100, \quad -12.25 < \phi < 5486.86,$$

$$M = 3839, \quad -75.89 < \phi < 1310795.75.$$

Also it was found that

$$(94) \quad \mathfrak{E}(\phi | \theta) = B \left\{ 1 + \frac{(1-2P_1)(1-2P_2)}{(1-P_1)(1-P_2)} (e^\theta - 1) + \frac{(M-1)P_1 P_2}{(1-P_1)(1-P_2)} (e^\theta - 1)^2 \right\}$$

where  $\mathfrak{E}(\phi | \theta)$  denotes the expected value of  $\phi$ , given the value of the parameter  $\theta$ . Thus when  $P_1 = P_2 = \frac{1}{4}$  we have  $B = \sqrt{\frac{2}{3}M}$  and so  $\mathfrak{E}(\phi | 0) = \sqrt{\frac{2}{3}M}$ . Hence when

$$M = 100, \quad \mathfrak{E}(\phi | 0) = 6.12372,$$

$$M = 3839, \quad \mathfrak{E}(\phi | 0) = 37.94239.$$

It is thus seen that the distribution of  $\phi$  might be represented by a Type III curve, since the distribution of  $\phi$  has a finite lower bound and a very long positive tail. In order to fit a Type III curve, we must know the second moment of the curve as well as its lower bound and mean. The general expression for the second moment about zero is too complicated to be printed and so only the numerical expressions obtained by giving special values to  $M$  are given below. These are:

$$(i) \quad M = 100$$

$$(95) \quad \mathfrak{E}(\phi^2 | \theta) = 112.41667 + 165.62963(e^\theta - 1) + 2493.33333(e^\theta - 1)^2 \\ + 1078.00000(e^\theta - 1)^3 + 4356.91667(e^\theta - 1)^4,$$

$$(ii) M = 3839$$

$$(96) \quad \xi(\phi^2 | \theta) = 4318.79213 + 6397.29625(e^\theta - 1) + 3684321.24073(e^\theta - 1)^2 \\ + 1636267.33255(e^\theta - 1)^3 + 261530062.11111(e^\theta - 1)^4.$$

Using the above results Type III curves were fitted to the distribution of  $\phi$ , and approximate values of the power functions  $\beta(w_M | \theta)$ , at level of significance .05, were calculated. This was obtained by evaluating  $P\{\phi > A_{.05} | \theta\}$  and assuming the distribution of  $\phi$  to be that given by the fitted curve. Then

$$(97) \quad \beta(w_M | \theta) = P\{\phi > A_{.05} | \theta\}.$$

The values obtained for the limiting and approximate power functions are given in Tables IIa, IIb. Unfortunately the agreement between the two is not satisfactory.

*Special Case.* For the cases  $P_1 = P_2 = \frac{1}{2}$  ( $M = 100, M = 400$ ) power functions were calculated on the assumption that for a given value of  $\theta$ , the random variable  $2M^{-1/2}(d_2 + d_3)$  is distributed normally about a mean  $M^{1/2}(e^\theta - 1)$  with standard deviation  $\sqrt{e^\theta(2 - e^\theta)}$ . This is approximately the case for the values of  $M$  considered. The approximate power functions so calculated are given in Tables IIIa, IIIb.

**7. Parabolic Test and  $\chi^2$  Test.** It is interesting to note the close connection between the parabolic test and the  $\chi^2$  test as introduced for intuitive reasons and normally used in testing for linkage. The  $\chi^2$  test consists of calculating the quantity

$$(98) \quad \chi^2 = \frac{1}{MP_1P_2(1 - P_1)(1 - P_2)} \{(1 - P_1)(1 - P_2)m_1 \\ - P_2(1 - P_1)m_2 - P_1(1 - P_2)m_3 + P_1P_2m_4\}^2$$

and rejecting the hypothesis  $H_0(\theta = 0)$  if  $|\chi| > a$  where

$$(99) \quad \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-t^2/2} dt = 1 - \alpha.$$

In the special case ( $P_1 = P_2 = \frac{1}{2}$ ) the parabolic test and the  $\chi^2$  test are identical; while comparing (98) and (79) we see that in the general case

$$(100) \quad u = \chi.$$

Hence in the general case the criterion used in the parabolic test may be written

$$(101) \quad \phi = v + B\chi^2.$$

(1) *Large Samples.* For large samples the first term of the expression  $v + B\chi^2$  is usually of small importance, since

$v$  is of form  $M^{-1} \times$  (linear function of the  $d_i$ 's), while  $B\chi^2$  is of form  $M^{-1} \times$  (quadratic function of the  $d_i$ 's).

For such samples the  $\chi^2$  test and parabolic test would appear to be nearly equivalent.

TABLE II  
*Limiting and Approximate Power Functions of Parabolic Test*

$$P_1 = P_2 = \frac{1}{2} \\ -\infty < \theta < 1.386$$

Table IIa $M = 100$			Table IIb $M = 3839$		
$\theta$	Power		$\theta$	Power	
	Limiting	Approximate		Limiting	Approximate
-2.00		0.90870	-0.25	0.99932	0.99853
-1.50	0.99880		-0.20	0.98502	0.97521
-1.40		0.77656	-0.15	0.87243	0.83620
-1.20	0.97915	0.69505	-0.10	0.54197	0.52066
-1.05	0.93786		-0.05	0.17827	0.19223
-1.00		0.58580	0.00	0.05000	0.04111
-0.90	0.85024		0.05	0.17827	0.21568
-0.75	0.70467	0.42755	0.10	0.54197	0.59517
-0.60	0.51532		0.15	0.87243	0.91641
-0.45	0.32258	0.21849	0.20	0.98502	0.99640
-0.30	0.16986	0.12504	0.25	0.99932	0.99999
-0.15	0.07905	0.05689			
-0.10	0.06280	0.04438			
-0.05	0.05318	0.03866			
0.00	0.05000	0.04069			
0.05	0.05318	0.05021			
0.10	0.06280	0.07429			
0.15	0.07905				
0.30	0.16986	0.26559			
0.45	0.32258				
0.60	0.51532	0.75854			
0.75	0.70467	0.94245			

THEOREM 4. *The limiting power function of the  $\chi^2$  test is*

$$(102) \quad \beta_\infty(w_\chi^2 | \vartheta) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+a} e^{-\frac{1}{2}[u - \vartheta(P_1 P_2)^{\frac{1}{2}}(1-P_1)^{-\frac{1}{2}}(1-P_2)^{-\frac{1}{2}}]^2} du$$

( $w_\chi$  denotes the region defined by the inequality  $|\chi| > a$ ).

This theorem may be proved by applying (46)-(48) to  $Q_\vartheta(x_1, x_2, x_3)$  in THEOREM 1 of the Appendix, and noting that  $u = \chi$  by (100).

We notice that  $\beta_{\infty}(w_{x^2} | \vartheta)$ , for a given value of  $\vartheta$ , has the same value for all values of  $M$ , unlike the limiting power function  $\beta_{\infty}(w_M | \vartheta)$  of the parabolic test. It is this point which accounts for the seeming paradox that, despite the manner in which the parabolic test was defined, for all values of  $\vartheta$  and  $M$

$$(103) \quad \beta_{\infty}(w_{x^2} | \vartheta) \geq \beta_{\infty}(w_M | \vartheta)$$

as may be deduced from (92) and (102). This does not mean that for any given  $\vartheta$  and all  $M$  sufficiently large the power function of the  $\chi^2$  test,  $\beta_M(w_{x^2} | \vartheta)$ ,

TABLE III  
*Approximate Power Function*

$$P_1 = P_2 = \frac{1}{2}$$

$$-\infty < \theta < 0.693$$

*Table IIIa.*  
 $M = 100$

$\theta$	Power
-0.45	0.96288
-0.40	0.92161
-0.35	0.85072
-0.30	0.74351
-0.25	0.60197
-0.20	0.44054
-0.15	0.28380
-0.10	0.15727
-0.05	0.07737
0.00	0.05000
0.05	0.08029
0.10	0.18177
0.15	0.36464
0.20	0.60278
0.25	0.82071
0.30	0.94975
0.35	0.99299

*Table IIIb.*  
 $M = 400$

$\theta$	Power
-0.25	0.99424
-0.20	0.95482
-0.15	0.79787
-0.10	0.47734
-0.05	0.16378
-0.02	0.06810
0.00	0.05000
0.02	0.06885
0.05	0.17609
0.10	0.55737
0.15	0.90213
0.20	0.99431
0.25	0.99995

is necessarily not less than the power function of the parabolic test,  $\beta_M(w_M | \vartheta)$ . For although, given any  $\epsilon > 0$ , there is a number  $M_{\epsilon, \vartheta}$  such that if  $M > M_{\epsilon, \vartheta}$

$$(104) \quad |\beta_M(w_{x^2} | \vartheta) - \beta_{\infty}(w_{x^2} | \vartheta)| < \epsilon$$

and

$$(105) \quad |\beta_M(w_M | \vartheta) - \beta_{\infty}(w_M | \vartheta)| < \epsilon$$

it may be that for such values of  $M_{\epsilon, \vartheta}$

$$(106) \quad 0 \leq \beta_{\infty}(w_{x^2} | \vartheta) - \beta_{\infty}(w_M | \vartheta) < 2\epsilon.$$

The above results show, however, how close the agreement between the power functions of the two tests is for large values of  $M$ . In fact we have

$$(107) \quad \lim_{M \rightarrow \infty} \beta_{\infty}(w_M | \vartheta) = \beta_{\infty}(w_{\chi^2} | \vartheta).$$

This may be easily proved, since as  $M$  increases  $w_M$  approximates to  $w_{\chi^2}$ .

(2) *Small Samples.* In order to obtain some idea of the relations between the two tests when  $M$  is small (i.e. less than 100), the case  $P_1 = P_2 = \frac{1}{4}$ ,  $M = 32$  was considered in some detail.

In this case our tests at 5% level of significance are respectively  $\chi^2$  test, reject if

$$(108) \quad |2y - z| > 8.315$$

parabolic test, reject if

$$(109) \quad (2y - z)^2 - \frac{3}{2}(2y + z) > 69.576$$

where

$$(110) \quad y = d_1 \quad z = d_2 + d_3.$$

All samples for which the verdicts of the two above tests would not agree were obtained. These were as follows:

(a) Samples for which  $H_0$  is accepted by  $\chi^2$  test, rejected by parabolic test

$y =$	1	0	-1	-2	Probability of drawing sample of this type when $H_0$ is true is 0.00320.
$z =$	-6	-8	-10	-12	

(b) Samples for which  $H_0$  is rejected by parabolic test, accepted by  $\chi^2$  test

$y =$	0	1	2	3	5	6	7	8	8	9	9	Probability of drawing sample of this type when $H_0$ is true is 0.00038.
	9	11	13	15	1	3	5	6	7	8	9	

Thus the probability of the two tests giving different verdicts when  $H_0$  is in fact true is only 0.00358.

It will be noted that the above results imply that

$$(111) \quad \beta_{32}(w_{32} | 0) - \beta_{32}(w_{\chi^2} | 0) = 0.00320 - 0.00038 = 0.00282;$$

i.e. that the true levels of significance of the two tests are not equal. This is to be expected, because of the discontinuity of the probability distribution of sample points, which makes it unlikely that the level of significance of either test is exactly .05.

Similarly we can obtain values of  $\beta_{32}(w_{32} | \theta) - \beta_{32}(w_{\chi^2} | \theta)$ , the differences in the powers of the two tests with respect to various alternative hypotheses. These values were obtained for a few values of  $\theta$ .

$\theta$	$\beta_{22}(w_{22}   \theta) - \beta_{22}(w_{x^2}   \theta)$
-0.5	0.01625
0.0	0.00282
0.5	-0.00006

These figures indicate that the parabolic test detects negative  $\theta$ 's better than the  $\chi^2$  test, but that the  $\chi^2$  test detects positive  $\theta$ 's better than the parabolic test, although the advantage in this latter case is minute.

The critical regions associated with the two tests may be represented by regions in the  $(y, z)$  plane. The critical region for the parabolic test will be defined by

$$(112) \quad (2y - z)^2 - \frac{2}{3}(2y + z) > \nu$$

and that for the  $\chi^2$  test,  $w_{x^2}$ , by

$$(113) \quad (2y - z)^2 > \nu'$$

where  $\nu \approx \nu'$ .

$w_{x^2}$  is therefore the complement of the region lying between the lines  $L_1, L_2$  with equations  $2y - z = \pm\sqrt{\nu'}$ ;  $w_M$  lies outside the parabola  $K$  with equation  $(2y - z)^2 - \frac{2}{3}(2y + z) = \nu$ .

Since  $\nu \approx \nu'$ ,  $K$  meets  $L_1, L_2$  at points near the respective intersections of  $L_1, L_2$  with the line  $2y + z = 0$ . See Figure 1.

In the diagram the regions  $V_1, V_2$  contain all sample points for which the  $\chi^2$  test rejects and the parabolic test accepts  $H_0$ ;  $U_1, U_2$  contain all sample points for which the  $\chi^2$  test accepts and the parabolic test rejects  $H_0$ .

For a given value of  $\theta$  it is known that the probability distribution is approximately such that the quantity

$$(114) \quad \psi_\theta^2 = \frac{\{y - \frac{1}{16}M(e^\theta - 1)\}^2}{\frac{1}{16}M + \frac{1}{16}M(e^\theta - 1)} + \frac{\{z + \frac{1}{8}M(e^\theta - 1)\}^2}{\frac{1}{16}M - \frac{1}{8}M(e^\theta - 1)} + \frac{\{y + z + \frac{1}{16}M(e^\theta - 1)\}^2}{\frac{1}{16}M + \frac{1}{16}M(e^\theta - 1)}$$

is distributed as  $\chi^2$  with 2 degrees of freedom.

The ellipses of equal density  $\psi_\theta^2 = \text{constant}$  have centers at points  $(\frac{1}{16}M[e^\theta - 1], -\frac{1}{8}M[e^\theta - 1])$  which must lie on the line  $2y + z = 0$ . When  $\theta = 0$  the center is at the origin, and the major and minor axes of the ellipse make angles of approximately  $99.5^\circ$  and  $9.5^\circ$  respectively with the  $y$ -axis. For small changes in  $\theta$  the angles of inclination of the major and minor axes of the ellipse to the coordinate axes are not greatly changed, and we see that as the center of the ellipse moves along the line  $2y + z = 0$  we have

(1)  $\theta$  increasing: center moves downwards, tending to increase  $P\{E \in U_2\} - P\{E \in V_2\}$  while  $P\{E \in V_1\}$  and  $P\{E \in U_1\}$  both become small. Thus  $\beta_M(w_M | \theta)$  tends to increase quicker than  $\beta_M(w_{x^2} | \theta)$ .



These are the proportions of results expected in the first three cells, if the hypothesis  $H_\theta$  specifying  $\theta$  be true.

Now, if  $H_\theta$  be true, we have

$$(117) \quad P\{n_1 = n'_1, n_2 = n'_2, n_3 = n'_3, n_4 = n'_4 \mid H_\theta\} \approx ce^{-\frac{1}{2}\chi_\theta^2}$$

where  $c$  is constant for a fixed sample size  $M$ , and

$$(118) \quad \frac{\chi_\theta^2}{M} = \sum_{i=1}^3 \frac{(n'_i - \theta^{n_i})^2}{\theta^{n_i}} + \frac{\left[ \sum_{i=1}^3 (n'_i - \theta^{n_i}) \right]^2}{1 - \sum_{i=1}^3 \theta^{n_i}}.$$

Hence the most frequent position(s) of the sample point  $E$  will be somewhere near the point  $T_\theta$ , which I shall therefore call the *center of density*. It will be noticed that, whatever be the value of  $\theta$ , the point  $T_\theta$  must lie on the line

$$n_1 - P_1P_2 = -[n_2 - P_1(1 - P_2)] = -[n_3 - P_2(1 - P_1)].$$

This line, a segment of which is the locus of the center of density for our set of admissible hypotheses, will be called the *line of density*.

In this space the parabolic test corresponds to a critical region comprising the exterior of a parabolic cylinder. The equation of the boundary of this critical region at level of significance .05 was found for the case  $P_1 = P_2 = \frac{1}{4}$ , and a model made of it. Also included in the model were the ellipsoids

$$(120) \quad \chi_\theta^2 = K_{.05}$$

where  $K_{.05}$  is a constant so chosen that

$$(121) \quad P\{\chi_\theta^2 > K_{.05} \mid \theta\} \approx .05$$

corresponding to

(i) the case when  $H_0$  is true

(ii) the cases when

$$(122) \quad (a) \ p_1 = \frac{3}{8}; p_2 = p_3 = \frac{5}{8}; p_4 = \frac{1}{8} \quad \text{i.e. } \theta = 0.41$$

$$(123) \quad (b) \ p_1 = \frac{1}{8}; p_2 = p_3 = \frac{7}{8}; p_4 = \frac{1}{8} \quad \text{i.e. } \theta = -0.69.$$

It was found that in the case  $P_1 = P_2 = \frac{1}{4}$  one axis of all the  $\chi_\theta^2$ -ellipsoids was perpendicular to the plane through the line of density and the axis of  $n_i$ . The generators of the boundary of the parabolic acceptance region are also perpendicular to this plane. (By "acceptance region" is meant the complement of the critical region. The acceptance region may be written symbolically  $\bar{w}_M$ .) There were further added to the model the intersections with this plane of the ellipsoids at probability level .01, corresponding to the three hypotheses considered above ( $\theta = 0, 0.41, -0.69$ ) and two others, viz.

$$(124) \quad p_1 = \frac{5}{8}; p_2 = p_3 = \frac{3}{8}; p_4 = \frac{2}{8} \quad \text{i.e. } \theta = 0.92,$$

$$(125) \quad p_1 = \frac{1}{8}; p_2 = p_3 = \frac{5}{8}; p_4 = \frac{3}{8} \quad \text{i.e. } \theta = -1.39.$$





of equal density include somewhat more of the acceptance region of the  $\chi^2_0$  test than of the parabolic acceptance region. This means that the  $\chi^2_0$  test would detect that the hypothesis  $H_0(\theta = 0)$  is false in these cases, less frequently than would the parabolic and  $\chi^2$  tests. We also notice that the center of density  $T_0$  leaves the parabolic acceptance region before it leaves the acceptance region of the  $\chi^2_0$  test as it moves along the line of density from the point where  $\theta = 0$ , whether the direction of motion of  $T_0$  corresponds to  $\theta$  increasing or decreasing. This also indicates that the  $\chi^2_0$  test would act less efficiently than the other two tests.

**9. Appendix.** In this appendix are obtained various results which, while essential to the main argument, would appear as digressions if they were interpolated as required. The numbering of equations in this appendix does not continue from that of the previous sections, but forms a separate group.

**LEMMA.** If  $f_0(m), f_1(m), \dots, f_s(m)$  be  $(s + 1)$  functions of the  $k$  variables  $m_1, m_2, \dots, m_k$  which are zero except for a finite number of sets of integral values of  $m_1, \dots, m_k$ ; and if  $w_0$  be a region in the space of  $m$ 's such that

$$(1) \quad f_0(m) \geq \sum_{i=1}^s a_i f_i(m) \quad \text{in } w_0$$

$$(2) \quad f_0(m) < \sum_{i=1}^s a_i f_i(m) \quad \text{in } \bar{w}_0$$

$a_1, a_2, \dots, a_s$  being arbitrary constants; then if  $w$  be any region such that

$$(3) \quad \sum_w f_i(m) = \sum_{w_0} f_i(m) \quad (i = 1, \dots, s),$$

we shall have

$$(4) \quad \sum_w f_0(m) \leq \sum_{w_0} f_0(m).$$

**PROOF.** Let

$$(5) \quad \begin{aligned} \delta &= \sum_{w_0} f_0(m) - \sum_w f_0(m) \\ &= \sum_{w_0 - ww_0} f_0(m) - \sum_{w - ww_0} f_0(m) \end{aligned}$$

where  $ww_0$  denotes the common part of  $w$  and  $w_0$ .

Hence the region  $w - ww_0$ , consisting of those points of  $w$  which are not in  $w_0$ , and so not in  $w_0$ , is contained in  $\bar{w}_0$ . Similarly the region  $w_0 - ww_0$  is contained in  $w_0$ . Hence, by inequalities (1),

$$(6) \quad \delta \geq \sum_{w_0 - ww_0} \left\{ \sum_{i=1}^s a_i f_i(m) \right\} - \sum_{w - ww_0} \left\{ \sum_{i=1}^s a_i f_i(m) \right\}$$

and so

$$(7) \quad \delta \geq \sum_{w_0} \left\{ \sum_{i=1}^s a_i f_i(m) \right\} - \sum_w \left\{ \sum_{i=1}^s a_i f_i(m) \right\}.$$

Since the total number of terms in each double summation is finite, we have

$$(8) \quad \delta \geq \sum_{i=1}^s a_i \left\{ \sum_{w_0} f_i(m) - \sum_w f_i(m) \right\}.$$

But

$$(9) \quad \sum_{w_0} f_i(m) = \sum_w f_i(m), \quad (i = 1, \dots, s).$$

Hence

$$\delta \geq 0, \quad \text{and} \quad \sum_w f_0(m) \leq \sum_{w_0} f_0(m).$$

A lemma similar to the lemma above, where the  $f$ 's are taken to be integrable functions and summation over the regions  $w$ ,  $w_0$  is replaced by integration over these regions, is given by Neyman and Pearson [9]. The proof given above follows the lines of the proof given in that paper.

**THEOREM 1.** *Suppose that, in a quadrinomial population:*

(i) *the cell probabilities are dependent on the number  $M$  of trials made, and are given by*

$$(10) \quad \begin{aligned} p_1 &= p_{01} + \varphi_M \\ p_2 &= p_{02} - \varphi_M \\ p_3 &= p_{03} - \varphi_M \\ p_4 &= p_{04} + \varphi_M \end{aligned}$$

where

$$(11) \quad \sum_{i=1}^4 p_{0i} = \sum_{i=1}^4 p_i = 1$$

and

$$(12) \quad \varphi_M = \lambda(e^{\vartheta M^{-\frac{1}{2}}} - 1)$$

(ii)

$$(13) \quad x_i = (m_i - Mp_{0i}) / (Mp_{0i})^{\frac{1}{2}} \quad (i = 1, 2, 3, 4)$$

where  $m_i$  = number of results falling in  $i$ -th cell.

(iii)  $w(x)$ , or briefly  $w$ , is a region in the space  $W$  of  $x_1, x_2, x_3$ ; and  $P_M(w)$  is the integral probability law of  $w$  corresponding to the values  $p_1, p_2, p_3, p_4$  of the cell probabilities given in (2) above when we have  $M$  independent trials.

Then

$$(14) \quad P_M(w) \rightarrow \frac{1}{(2\pi)^{\frac{1}{2}} p_{04}^{\frac{1}{2}}} \int \int \int e^{-\frac{1}{2} Q_{\theta}(x_1, x_2, x_3)} dx_1 dx_2 dx_3$$

uniformly over  $W$  as  $M \rightarrow \infty$ , where

$$\begin{aligned}
 Q_{\vartheta}(x_1, x_2, x_3) = & \sum_{i=1}^3 x_i^2(1 + p_{0i}p_{04}^{-1}) + 2p_{04}^{-1} \sum_{i < j \leq 3} x_i x_j (p_{0i}p_{0j})^{\frac{1}{2}} \\
 (15) \quad & - 2\lambda\vartheta \{x_1(p_{01}^{-1} - p_{02}p_{04}^{-1}) - x_2(p_{02}^{-1} + p_{03}p_{04}^{-1}) \\
 & - x_3(p_{03}^{-1} + p_{03}p_{04}^{-1})\} + \lambda^2\vartheta^2 \sum_{i=1}^4 p_{0i}^{-1}.
 \end{aligned}$$

This theorem may be proved by the same method as that used by F. N. David [2] in proving the generalized theorem of Laplace.

I would like to thank Professor Neyman for his invaluable suggestions and advice in the preparation of this paper.

#### REFERENCES

- [1] JOSEPH F. DALY, *Annals of Math. Stat.*, Vol. 11 (1940), p. 1.
- [2] F. N. DAVID, *Stat. Res. Mem.*, Vol. 2 (1938), p. 69.
- [3] R. A. FISHER, *Statistical Methods for Research Workers*, 6th. ed. (1936), Oliver and Boyd, London.
- [4] J. NEYMAN, *Phil. Trans., A*, Vol. 236 (1937), p. 333.
- [5] J. NEYMAN, *Skand. Actuar. Tidskr.*, Vol. 20 (1937), p. 149.
- [6] J. NEYMAN, *Ann. Math. Stat.*, Vol. 9 (1938), p. 69.
- [7] J. NEYMAN AND E. S. PEARSON, *Biometrika*, Vol. 20A (1928), p. 223.
- [8] J. NEYMAN AND E. S. PEARSON, *Phil. Trans., A*, Vol. 231 (1933), p. 289.
- [9] J. NEYMAN AND E. S. PEARSON, *Stat. Res. Mem.*, Vol. 1 (1936), p. 1.
- [10] J. NEYMAN AND E. S. PEARSON, *Stat. Res. Mem.*, Vol. 2 (1938), p. 25.

DEPARTMENT OF STATISTICS,  
UNIVERSITY COLLEGE,  
LONDON, ENGLAND

# REDUCTION OF A CERTAIN CLASS OF COMPOSITE STATISTICAL HYPOTHESES

BY GEORGE W. BROWN

**1. Introduction.** A situation frequently met in sampling theory is the following:  $x$  has distribution  $f(x, \theta)$ , where  $\theta$  is an unknown parameter, and for samples  $(x_1, \dots, x_n)$  there exists in the sample space  $E_n$  a family of  $(n - 1)$ -dimensional manifolds upon each of which the distribution is independent of  $\theta$ ; in addition there is a residual one-dimensional manifold available for estimating  $\theta$ . For example, suppose there exists a sufficient statistic  $T$  for  $\theta$ , then on the manifolds  $T = T_0$  there is defined an induced distribution which is independent of the parameter.

A similar situation is observed when  $\theta$  is a "location" or "scale" parameter. Let  $x$  have the distribution  $f(x - a)$  for some  $a$ , then the set  $(x_2 - x_1, x_3 - x_1, \dots, x_n - x_1)$ , or any equivalent set, such as  $(x_2 - \bar{x}, \dots, x_n - \bar{x})$ , have a joint distribution independent of  $a$ , and there is a residual distribution corresponding to each particular configuration  $(x_2 - x_1, \dots, x_n - x_1)$ . Fisher [1] and Pitman [5] have examined the residual distributions in connection with the problem of estimating scale and location parameters. In this paper we shall be concerned primarily, not with the residual distribution, but with the remainder of the sample information, corresponding to the  $(n - 1)$ -dimensional distribution which is independent of the parameter. It is found, in a rather broad class of distributions, that the part of the sample *not* used for estimation *determines, except for the parameter value, the original functional form of the distribution of  $x$ .*

This paper is devoted mainly to a study of particular classes of distributions having the property mentioned above. We consider also the theoretical application of this property to certain types of *composite* hypotheses which may be reduced thereby to equivalent *simple* hypotheses.<sup>1</sup> The principal results of this nature may be summed up as follows: If  $x$  has distribution of the form  $f(x, \theta)$ , where  $\theta$  is either a location or scale parameter, or a vector denoting both, then there exists, in samples  $(x_1, \dots, x_n)$  a set of functions  $y_i(x_1, \dots, x_n)$ ,  $i = 1, 2, \dots, p$ ,  $p < n$ , having joint distribution  $D(y_1, \dots, y_p)$  independent of  $\theta$ , and such that the converse statement holds, namely, *if  $\{y_i\}$  have the distribution  $D(y_1, \dots, y_p)$ , then  $x$  has, for some  $\theta$ , a distribution of the form  $f(x, \theta)$ .* There is a corresponding statement when  $x$  has a distribution of the form  $f(x - \sum a_i u_i)$ , where the  $\{a_i\}$  are parameters, and the  $\{u_i\}$  are regression variables.

<sup>1</sup> We use the terms simple and composite hypotheses in the sense of Neyman and Pearson [2].

**2. Location and Scale.** This section is devoted to the study of functions of the sample observations which are such that their distributions determine the distribution of  $x$ , except possibly for location and scale.

It will be assumed that associated with  $x$  there is a function  $F(x)$  such that

- (a)  $F(x)$  is monotone non-decreasing,  
 (b)  $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$ , and (c)  $F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$

with the normalization  $F(x)$  upper semi-continuous.  $F(x)$  is the probability that the random variate takes a value less than or equal to  $x$ . If  $F(x)$  is associated with the random variate  $x$  we say that  $x$  has the distribution  $F(x)$ . If  $g(x)$  is a Borel-measurable function, the Lebesgue-Stieltjes integral  $\int_{-\infty}^{\infty} g(x) dF(x)$  is denoted by  $E[g(x)]$ .

The characteristic function  $\varphi(t) = E(e^{itx})$  determines  $F(x)$ , that is, if  $\int_{-\infty}^{\infty} e^{itx} dG(x) = \int_{-\infty}^{\infty} e^{itx} dF(x)$ , then  $F(x) = G(x)$ .

Similarly, let  $F(x_1, \dots, x_k)$  be such that

- (a)  $F(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_k) \geq F(x_1, \dots, x_i, \dots, x_k)$  for  $h > 0$  and  $i = 1, 2, \dots, k$ ;  
 (b)  $\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_k) = 0, i = 1, 2, \dots, k$ ;  
 (c)  $\lim_{x_1, \dots, x_k \rightarrow \infty} F(x_1, \dots, x_k) = 1$ ;

with the normalization  $F(x_1, \dots, x_k)$  continuous on the right in each  $x_i$ . If  $F(x_1, \dots, x_k)$  is associated with  $x_1, \dots, x_k$  we say that  $x_1, \dots, x_k$  have the joint distribution  $F(x_1, \dots, x_k)$ . As before,  $E[H(x_1, \dots, x_k)] = \int_{R_k} H dF$ , where  $R_k$  is the Euclidean  $k$ -space. It is well known that under such conditions, given Borel-measurable functions  $y_i(x_1, \dots, x_k), i = 1, \dots, p, p \leq k$ , then  $G(y_1, \dots, y_p) = \int_{R(y)} dF(x_1, \dots, x_k)$ , where  $R(y)$  is the region  $[y_1(x_1, \dots, x_k) \leq y_1, \dots, y_p(x_1, \dots, x_k) \leq y_p]$ , is again a distribution function satisfying the conditions above. Moreover,  $\int_R g(y_1, \dots, y_p) dG(y_1, \dots, y_p) = \int_{R'} g[y_1(x_1, \dots, x_k), \dots, y_p(x_1, \dots, x_k)] dF$ , where  $R'$  is the set of all points  $(x_1, \dots, x_k)$  such that  $[y_1(x_1, \dots, x_k), \dots, y_p(x_1, \dots, x_k)] \in R$ .

If  $x$  has distribution  $F(x)$ , then, by definition, the set  $(x_1, \dots, x_n)$  is a sample from this distribution if  $x_1, \dots, x_n$  have the joint distribution  $F(x_1) \dots F(x_n)$ .

The following theorem states that two distributions giving rise, in sampling, to the same distribution of the set  $x_1 - x_n, x_2 - x_n, \dots, x_{n-1} - x_n$ , with  $n \geq 3$ , can differ at most by a translation, that is, the distribution of that set determines the original distribution except for location.

**THEOREM 1A:** Let  $x$  have the distribution  $F(x)$ . Denote by  $S$  the set of zeros of

$\int e^{itz} dF(x)$  and denote by  $\epsilon$  the g.l.b. of  $|t|$  for  $t$  in  $S$ . Suppose that the complement of  $S$  is  $\epsilon$ -connected.<sup>2</sup> Suppose that  $x'$  has distribution  $G(x')$ , and let  $x_1, \dots, x_n$  and  $x'_1, \dots, x'_n$  be samples. Then the set  $w_\alpha = x_\alpha - x_n$ ,  $\alpha = 1, \dots, n-1$ , have the same joint distribution as the set  $w'_\alpha = x'_\alpha - x'_n$  if and only if there exists a constant  $a$  such that  $x' + a$  and  $x$  have the same distribution.

PROOF: The sufficiency of the condition follows immediately, since  $w'_\alpha = x'_\alpha - x'_n = (x'_\alpha + a) - (x'_n + a)$ .

In establishing necessity, only the fact that  $w_1, w_2$  have the same joint distribution as  $w'_1, w'_2$  is needed. This hypothesis implies that

$$E\{e^{it_1 w_1 + t_2 w_2}\} = E\{e^{it_1 w'_1 + t_2 w'_2}\},$$

that is,

$$E\{e^{it_1(x_1 - x_n) + t_2(x_2 - x_n)}\} = E\{e^{it_1(x'_1 - x'_n) + t_2(x'_2 - x'_n)}\}.$$

Set  $\varphi(t) = E(e^{itz})$ ,  $\psi(t) = E(e^{itz'})$ . The relation above becomes

$$(1) \quad \varphi(t_1)\varphi(t_2)\varphi(-t_1 - t_2) = \psi(t_1)\psi(t_2)\psi(-t_1 - t_2).$$

Consider equation (1) for values of  $t_1, t_2$  in the neighborhood of  $t = 0$ .  $\varphi(0) = \psi(0) = 1$ , hence there is an interval  $|t| < \delta$ , in which  $\varphi(t)$  and  $\psi(t)$  do not vanish. It is easily shown that  $\varphi(t)$  and  $\psi(t)$  are each continuous, since  $e^{itz}$ , in the neighborhood of  $t = 0$ , is continuous uniformly for any bounded interval of  $x$ , and since  $A$  may be chosen so that  $1 - F(A)$  and  $F(-A)$  are both as small as desired. In the interval  $|t| < \delta$  the function  $f(t) = \varphi(t)/\psi(t)$  is continuous. Also,  $\varphi(-t) = \overline{\varphi(t)}$  and  $\psi(-t) = \overline{\psi(t)}$ . Setting  $t_2 = 0$  in (1) we obtain  $\varphi(t)\varphi(-t) = \psi(t)\psi(-t)$ , hence  $|\varphi(t)| = |\psi(t)|$ , that is,  $|f(t)| = 1$ .  $f(t)$  takes values on the unit circle of the complex plane, and  $f(0) = 1$ , hence there is an interval  $|t| < \delta'$  such that  $z = f(t)$  lies on an arc  $\gamma$ , of length less than  $2\pi$ , containing the point  $z = 1$ . Now consider the functional equation (1) for  $|t_1| < \frac{1}{2}\delta'$ ,  $|t_2| < \frac{1}{2}\delta'$ . (1) becomes

$$f(t_1)f(t_2)f(-t_1 - t_2) = 1.$$

The interval  $|t| < \delta'$  was so chosen that for  $|t_1| < \frac{1}{2}\delta'$ ,  $|t_2| < \frac{1}{2}\delta'$ , it is possible to define a single-valued branch of the argument of  $f(t_1)$ ,  $f(t_2)$ , and  $f(t_1 + t_2)$ . Letting  $t_2 = 0$  we have  $f(t)f(-t) = 1$ , hence, replacing  $f(-t_1 - t_2)$  by  $1/f(t_1 + t_2)$  in the last equation, we have

$$f(t_1)f(t_2) = f(t_1 + t_2).$$

$\arg f(t_1)$ ,  $\arg f(t_2)$ , and  $\arg f(t_1 + t_2)$  are uniquely determined, except for some fixed multiple of  $2\pi$ . If we choose the principal value of the argument, i.e., so

<sup>2</sup> The set  $S$  is  $\epsilon$ -connected if any two points  $p, q$ , in  $S$  can be connected by an  $\epsilon$ -chain, i.e., there exists a set  $p_0 = p, p_1, \dots, p_{n-1}, p_n = q$ , such that  $|p_i - p_{i-1}| < \epsilon$ ,  $i = 1, 2, \dots, n$ .

that  $0 \leq \arg f(t) < 2\pi$ , we must have

$$\arg f(t_1) + \arg f(t_2) = \arg f(t_1 + t_2)$$

for  $|t_1| < \frac{1}{2}\delta'$ ,  $|t_2| < \frac{1}{2}\delta'$ . Since  $\arg f(t)$  is continuous, any solution of this well known functional equation must be of the form  $\arg f(t) = at$ .  $|f(t)| = 1$ , therefore there exists a constant  $a$  such that  $f(t) = e^{iat}$ , for  $|t| < \frac{1}{2}\delta'$ , that is,  $\varphi(t) = e^{iat}\psi(t)$ , for  $|t| < \frac{1}{2}\delta'$ . By use of (1) this may be extended to hold for all  $t$  such that  $|t| < \epsilon$ , where  $\epsilon$  is the minimum modulus of all  $t$  such that  $\varphi(t) = 0$ . (1) may now be used to extend the relation for all  $t$  such that  $\varphi(t) \neq 0$  by choosing an  $\epsilon$ -chain connecting the origin to the point  $t$ . We know already that  $\varphi(t) = e^{iat}\psi(t)$  if  $\varphi(t) \neq 0$ , hence it holds for all  $t$ . This relation says that  $E(e^{itz}) = E(e^{it(x'+a)})$ , hence  $x' + a$  and  $x$  have the same distribution, thus completing the demonstration of the theorem.

It should be remarked that the set  $(x_1 - x_n, \dots, x_{n-1} - x_n)$  may be replaced in Theorem Ia by any equivalent set, for example,  $(x_1 - \bar{x}, \dots, x_{n-1} - \bar{x})$ .

The next result is of the same nature as Theorem Ia except for the replacement of the location parameter by a scale (positive or negative) parameter.

**THEOREM IB:** *Let  $x$  have distribution  $F(x)$ , such that the zeros of  $\int_{-\infty}^{\infty} e^{it(\log|x|)} dF(x)$  are nowhere dense, and let  $x'$  have distribution  $G(x')$ . Let  $x_1, \dots, x_n$  and  $x'_1, \dots, x'_n$  be samples from the distributions of  $x$  and  $x'$ , with  $n \geq 3$ , then the set  $w_\alpha = x_\alpha/x_n$ ,  $\alpha = 1, \dots, n-1$ , have the same distribution as the set  $w'_\alpha = x'_\alpha/x'_n$  if and only if there exists a constant  $c$  such that  $cx'$  and  $x$  have the same distribution.*

**PROOF:** The sufficiency of the condition is evident. Suppose, then, as before, that  $w_1, w_2$  have the same joint distribution as  $w'_1, w'_2$ .  $\log|w_1|$  and  $\log|w_2|$  have the same joint distribution as  $\log|w'_1|$  and  $\log|w'_2|$ , hence by application of Theorem Ia to  $\log|x|$  and  $\log|x'|$  it follows (since the complement of a nowhere dense set is  $\epsilon$ -connected for every  $\epsilon$ ) that there exists a constant  $a$  such that

$$\int_{-\infty}^{\infty} e^{it \log|x|} dF(x) = \int_{-\infty}^{\infty} e^{it [\log|x'| - a]} dG(x).$$

Let  $y = e^{-a}x'$ , then  $|x|$  and  $|y|$  have the same distribution, and

$$(2) \quad \int e^{it \log|x|} dF(x) = \int e^{it \log|y|} dH(y),$$

where  $y$  has distribution  $H(y)$ . We now have to show that either  $y$  or  $-y$  has the distribution of  $x$ , that is, it must be shown that either  $H(y) = F(y)$ , or  $H(y) = 1 - F(-y)$ .

By the first part of the theorem the functions  $u_1 = y_1/y_2$  and  $u_2 = y_2/y_3$  have the same joint distribution as  $w_1, w_2$ . It is clear that the mean value of any function of  $u_1$  and  $u_2$  is the same as the mean value of the corresponding func-



tion of  $w_1$  and  $w_2$ . Hence

$$\iiint e^{i[t_1 \log |w_1| + t_2 \log |w_2|]} \operatorname{sgn} w_1 \operatorname{sgn} w_2 dF(x_1) dF(x_2) dF(x_3) \\ \iiint e^{i[t_1 \log |u_1| + t_2 \log |u_2|]} \operatorname{sgn} u_1 \operatorname{sgn} u_2 dH(y_1) dH(y_2) dH(y_3),$$

where  $\operatorname{sgn} x = 1$ , for  $x \geq 0$ ,  $\operatorname{sgn} x = -1$  for  $x < 0$ .

$$(\operatorname{sgn} w_1)(\operatorname{sgn} w_2) = (\operatorname{sgn} x_1)(\operatorname{sgn} x_2),$$

so that the last equation becomes

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i[t_1 (\log |x_1| - \log |x_2|) + t_2 (\log |x_2| - \log |x_3|)]} \operatorname{sgn} x_1 \operatorname{sgn} x_2 dF(x_1) dF(x_2) dF(x_3) \\ (3) \quad & = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i[t_1 (\log |y_1| - \log |y_2|) + t_2 (\log |y_2| - \log |y_3|)]} \operatorname{sgn} y_1 \\ & \quad \times \operatorname{sgn} y_2 dH(y_1) dH(y_2) dH(y_3). \end{aligned}$$

Set

$$\begin{aligned} \psi_1(t) &= \int e^{it \log |x|} dF(x); & \varphi_1(t) &= \int e^{it \log |y|} dH(y) \\ \psi_2(t) &= \int e^{it \log |x|} \operatorname{sgn} x dF(x); & \varphi_2(t) &= \int e^{it \log |y|} \operatorname{sgn} y dH(y). \end{aligned}$$

From (3) we have  $\psi_2(t_1)\psi_2(t_2)\psi_1(-t_1-t_2) = \varphi_2(t_1)\varphi_2(t_2)\varphi_1(-t_1-t_2)$  for all  $t_1, t_2$ , and from (2) we have  $\psi_1(t) = \varphi_1(t)$  for all  $t$ , hence, if  $\psi_1(-t_1-t_2) \neq 0$ ,  $\psi_2(t_1)\psi_2(t_2) = \varphi_2(t_1)\varphi_2(t_2)$ . By hypothesis the zeros of  $\psi_1(t)$  are nowhere dense, hence if  $\psi_1(-t_1-t_2) = 0$  there is a sequence  $t^{(n)}$ , such that  $t^{(n)} \rightarrow -t_1-t_2$  and  $\psi_1(t^{(n)}) \neq 0$ . Now take an arbitrary sequence  $t_1^{(n)}$  such that  $t_1^{(n)} \rightarrow t_1$ , then  $t_2^{(n)} = -t^{(n)} - t_1^{(n)}$  must tend to  $t_2$ . For each  $n$  we have  $\psi_2(t_1^{(n)})\psi_2(t_2^{(n)}) = \varphi_2(t_1^{(n)})\varphi_2(t_2^{(n)})$ . All the functions appearing are continuous, thus we see that  $\psi_2(t_1)\psi_2(t_2) = \varphi_2(t_1)\varphi_2(t_2)$  for all  $t_1, t_2$ . From this it follows directly that either  $\psi_2(t) = \varphi_2(t)$  for all  $t$  or  $\psi_2(t) = -\varphi_2(t)$  for all  $t$ . We have<sup>3</sup>

$$\begin{aligned} \psi_1(t) &= \int_0^{\infty} e^{it \log x} dF(x) + \int_{-\infty}^0 e^{it \log (-x)} dF(x) \\ \psi_2(t) &= \int_0^{\infty} e^{it \log x} dF(x) - \int_{-\infty}^0 e^{it \log (-x)} dF(x) \end{aligned}$$

<sup>3</sup> The assumption has been made implicitly that  $F(x)$  and  $G(x)$  are continuous at  $x = 0$ , otherwise the distribution of  $x_i/x_n$  is not properly defined, and the functions  $\varphi_i(t)$  and  $\psi_i(t)$  are then not defined. Similar assumptions will be made whenever necessary in later theorems.

$$\varphi_1(t) = \int_0^\infty e^{it \log x} dH(x) + \int_{-\infty}^0 e^{it \log(-x)} dH(x)$$

$$\text{and} \quad \varphi_2(t) = \int_0^\infty e^{it \log x} dH(x) - \int_{-\infty}^0 e^{it \log(-x)} dH(x).$$

Combining these expressions with the relations obtained above leads, by Fourier inversion, to the result that either  $F(x) = H(x)$  or  $H(x) = 1 - F(-x)$ . We have shown that either  $y$  or  $-y$  has the same distribution as  $x$ , that is, either  $e^{-a}x'$  or  $-e^{-a}x'$  has the same distribution as  $x$ .

Theorem Ib states essentially that the joint distribution of the set  $x_\alpha/x_n$ ,  $\alpha = 1, \dots, n-1$ , determines the distribution of  $x$  except for a scale parameter and possibly a reflection. In the event that  $x$  has an asymmetrical distribution, and if it is desired to rule out negative changes of scale, a variation of this procedure is necessary. The next result is appropriate for this situation.

**THEOREM IC:** *Let  $x$  have distribution  $F(x)$  such that the zeros of  $\int e^{it \log |x|} dF(x)$  are nowhere dense, and let  $x'$  have distribution  $G(x')$ . Let  $x_1, \dots, x_n$  and  $x'_1, \dots, x'_n$  be samples from the distributions of  $x$  and  $x'$ , with  $n \geq 3$ . Express  $x_1, \dots, x_n$  and  $x'_1, \dots, x'_n$  in spherical coordinates*

$$\begin{aligned} x_1 &= r \cos \theta_1, & x'_1 &= r' \cos \theta'_1 \\ x_2 &= r \sin \theta_1 \cos \theta_2, & x'_2 &= r' \sin \theta'_1 \cos \theta'_2 \\ &\vdots & &\vdots \\ x_n &= r \sin \theta_1 \sin \theta_2 \dots \sin \theta_{n-1}, & x'_n &= r' \sin \theta'_1 \sin \theta'_2 \dots \sin \theta'_{n-1}. \end{aligned}$$

*Then  $\theta_1, \dots, \theta_{n-1}$  have the same joint distribution as  $\theta'_1, \dots, \theta'_{n-1}$  if and only if there exists a positive constant  $k$  such that  $kx'$  and  $x$  have the same distribution.*

**PROOF:** Sufficiency of the condition is an immediate consequence of the fact that  $\theta_1, \dots, \theta_{n-1}$  are invariant under the transformation  $x = kx'$ , with  $k > 0$ . If  $\theta_1, \dots, \theta_{n-1}$  have the same joint distribution as  $\theta'_1, \dots, \theta'_{n-1}$  then the set  $\{x_\alpha/x_n\}$  have the same joint distribution as the set  $\{x'_\alpha/x'_n\}$ , hence, by Theorem Ib, there exists a constant  $c$  such that  $cx'$  has the same distribution as  $x$ . To establish necessity of the condition we must show that  $|c|x'$  has the same distribution as  $x$ .

Set  $y = |c|x'$ , and let  $y_1, \dots, y_n$  be expressed in spherical coordinates;  $y_1, \dots, y_n$  have the same angular coordinates  $\theta'_1, \dots, \theta'_{n-1}$ . This implies that  $x_1/r$  and  $x_2/r$  have the same joint distribution as  $y_1/R$  and  $y_2/R$ , where  $R = \sqrt{y_1^2 + \dots + y_n^2}$ ;  $\frac{x_1}{r} = x_1/|x_2|$ , therefore  $x_1/|x_2|$  has the same distribution as  $y_1/|y_2|$ , so that

$$\iint e^{it \log \left| \frac{x_1}{x_2} \right|} \operatorname{sgn} \left( \frac{x_1}{x_2} \right) dF(x_1) dF(x_2) = \iint e^{it \log \left| \frac{y_1}{y_2} \right|} \operatorname{sgn} \left( \frac{y_1}{y_2} \right) dH(y_1) dH(y_2)$$

if  $y$  has distribution  $H(y)$ .  $\text{Sgn} \left( \frac{x_1}{|x_2|} \right) = \text{sgn } x_1$ , so that the last equation yields

$$\begin{aligned} \int_{-\infty}^{\infty} e^{it \log |x|} \text{sgn } x dF(x) \cdot \int_{-\infty}^{\infty} e^{-it \log |x|} dF(x) \\ = \int_{-\infty}^{\infty} e^{it \log |x|} \text{sgn } x dH(x) \cdot \int_{-\infty}^{\infty} e^{-it \log |x|} dH(x). \end{aligned}$$

We know already that  $|x|$  and  $|y|$  have the same distribution, so that

$$(4) \quad \int_{-\infty}^{\infty} e^{it \log |x|} dF(x) = \int_{-\infty}^{\infty} e^{it \log |x|} dH(x),$$

thus

$$(5) \quad \int_{-\infty}^{\infty} e^{it \log |x|} \text{sgn } x dF(x) = \int_{-\infty}^{\infty} e^{it \log |x|} \text{sgn } x dH(x),$$

except possibly for zeros of  $\int_{-\infty}^{\infty} e^{-it \log |x|} dF(x)$ . By hypothesis the exceptional points are nowhere dense, so that, by continuity, (5) holds for all  $t$ . (4) and (5) together imply, as in the proof of Theorem Ib, that  $F(x) \equiv H(x)$ , i.e.,  $x$  and  $|c| x'$  have the same distribution.

The next three results are generalizations of Theorems Ia, b, c, to analogous multivariate situations. The first of these is a direct generalization of Theorem Ia.

**THEOREM IIA:** Let  $x_1, \dots, x_k$  have joint distribution  $F(x_1, \dots, x_k)$  such that the complement of the set  $S$  of zeros of  $\int e^{i \sum_{r=1}^k t_r x_r} dF(x_1, \dots, x_k)$  is  $\epsilon$ -connected, where  $\epsilon$  is the g.l.b. of  $|t|$  for  $(t)$  in  $S$ , and let  $y_1, \dots, y_k$  have joint distribution  $G(y_1, \dots, y_k)$ . Let  $(x_1^\alpha, \dots, x_k^\alpha)$  and  $(y_1^\alpha, \dots, y_k^\alpha)$ ,  $\alpha = 1, \dots, n$ , be samples from these distributions, with  $n \geq 3$ . Then  $w_{i\beta} = x_i^\beta - x_i^\alpha$ ,  $i = 1, \dots, k$ ,  $\beta = 1, \dots, n-1$ , have the same joint distribution as the corresponding set  $v_{i\beta} = y_i^\beta - y_i^\alpha$  if and only if there exist constants  $a_1, \dots, a_k$  such that  $y_1 + a_1, \dots, y_k + a_k$  have the same joint distribution as  $x_1, \dots, x_k$ .

**PROOF:** Set

$$\varphi(t_1, \dots, t_k) = \int e^{i \sum_{r=1}^k t_r x_r} dF(x_1, \dots, x_k),$$

$$\psi(t_1, \dots, t_k) = \int e^{i \sum_{r=1}^k t_r y_r} dG(y_1, \dots, y_k).$$

If  $w_{i\beta}$ ,  $i = 1, \dots, k$ ,  $\beta = 1, 2$ , have the same joint distribution as  $v_{i\beta}$ , then, as in the proof of Theorem Ia, we have

$$\begin{aligned} (6) \quad & \varphi(t_{11}, \dots, t_{k1}) \varphi(t_{12}, \dots, t_{k2}) \varphi(-t_{11} - t_{12}, \dots, -t_{k1} - t_{k2}) \\ & = \psi(t_{11}, \dots, t_{k1}) \psi(t_{12}, \dots, t_{k2}) \psi(-t_{11} - t_{12}, \dots, -t_{k1} - t_{k2}). \end{aligned}$$

Again, as before,  $|\varphi| = |\psi|$ ;  $\varphi(t_1, \dots, t_k)$  and  $\psi(t_1, \dots, t_k)$  are continuous;  $\varphi(0, 0, \dots, 0) = \psi(0, 0, \dots, 0) = 1$ . There will exist a neighborhood  $N$  of  $(0, 0, \dots, 0)$  such that for  $(t_1, \dots, t_k) \in N$  the function  $f(t_1, \dots, t_k) = \frac{\varphi(t_1, \dots, t_k)}{\psi(t_1, \dots, t_k)}$  is defined and continuous. Then there will exist a neighborhood  $N' \subset N$  such that in  $N'$  there exists a uniquely determined branch of  $\arg f(t_1, \dots, t_k)$ , continuous in  $N'$ , and such that if  $(t_1, \dots, t_k) \in N'$  and  $(u_1, \dots, u_k) \in N'$  then  $\arg f(t_1 + u_1, \dots, t_k + u_k)$  is also uniquely determined and continuous. For  $(t) \in N'$  and  $(u) \in N'$ ,  $\arg f$  satisfies the relation

$$\arg f(t_1, \dots, t_k) + \arg f(u_1, \dots, u_k) = \arg f(t_1 + u_1, \dots, t_k + u_k).$$

It is easily shown that any continuous function satisfying the equation above must be of the form  $\Sigma a_r t_r$ , therefore

$$(7) \quad \varphi(t_1, \dots, t_k) = e^{i \sum a_r t_r} \psi(t_1, \dots, t_k); \quad (t) \in N'.$$

Just as in the proof of Ia the relation (7) may be extended, by use of (6), to hold for all  $t$ . This implies, finally, that the set  $\{y_i + a_i\}$  have the same joint distribution as the set  $\{x_i\}$ .

Theorem IIb is a generalization of Theorem Ib to multivariate distributions.

**THEOREM IIb:** Let  $x_1, \dots, x_k$  have distribution  $F(x_1, \dots, x_k)$  such that the zeros of  $\int e^{i \Sigma t_r \log |x_r|} dF(x_1, \dots, x_k)$  are nowhere dense, and let  $y_1, \dots, y_k$  have distribution  $G(y_1, \dots, y_k)$ . Let  $(x_1^\alpha, \dots, x_k^\alpha)$  and  $(y_1^\alpha, \dots, y_k^\alpha)$ ,  $\alpha = 1, \dots, n$ , be samples, with  $n \geq 3$ . Then the set  $w_{i\beta} = x_i^\beta / x_i^\alpha$ ,  $i = 1, \dots, k$ ,  $\beta = 1, \dots, n - 1$ , have the same joint distribution as the corresponding set  $v_{i\beta} = y_i^\beta / y_i^\alpha$  if and only if there exist constants  $c_1, \dots, c_k$  such that the set  $c_i y_i$  have the same distribution as the  $x_i$ .

**PROOF:** The demonstration is parallel to that of Theorem Ib. By Theorem IIa there exist  $a_1, \dots, a_k$  such that

$$E(e^{i \Sigma t_r \log |x_r|}) = E(e^{i \Sigma t_r (\log |y_r| + a_r)}).$$

Set  $z_r = e^{a_r} y_r$ , then

$$(8) \quad \int e^{i \Sigma t_r \log |x_r|} dF(x_1, \dots, x_k) = \int e^{i \Sigma t_r \log |z_r|} dH(z_1, \dots, z_k),$$

where  $(z_1, \dots, z_k)$  have distribution function  $H(z_1, \dots, z_k)$ .

We shall continue the proof from here under the assumption that  $k = 2$ . It will be evident how the proof goes for any  $k$ . We have, since  $x_r^\beta / z_i^\beta$  have the same joint distribution as  $x_r^\beta / x_i^\beta$ ,

$$(9) \quad \begin{aligned} & \int \int \int_{-\infty}^{\infty} e^{i \Sigma t_r \beta (\log |x_r^\beta| - \log |x_r^\beta|)} \operatorname{sgn} \left( \frac{x_1^1}{x_1^3} \right) \operatorname{sgn} \left( \frac{x_1^2}{x_1^3} \right) dF(x_1^1, x_1^2) dF(x_1^2, x_1^3) dF(x_1^3, x_1^3) \\ &= \int \int \int e^{i \Sigma t_r \beta (\log |x_r^\beta| - \log |x_r^\beta|)} \operatorname{sgn} \left( \frac{x_1^1}{x_1^3} \right) \operatorname{sgn} \left( \frac{x_1^2}{x_1^3} \right) dH(x_1^1, x_1^2) dH(x_1^2, x_1^3) dH(x_1^3, x_1^3). \end{aligned}$$

Both members of (9) are evaluated as products, just as was done in previous proofs, and from the result, combined with (8), we conclude, as in Theorem Ib, that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\mathbf{z}t_r \log |x_r|} \operatorname{sgn} x_1 dF(x_1, x_2) = s_1 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\mathbf{z}t_r \log |x_r|} \operatorname{sgn} x_1 dH(x_1, x_2),$$

where  $s_1 = \pm 1$ , for all  $(t_1, t_2)$ . Similarly

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\mathbf{z}t_r \log |x_r|} \operatorname{sgn} x_2 dF(x_1, x_2) = s_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\mathbf{z}t_r \log |x_r|} \operatorname{sgn} x_2 dH(x_1, x_2)$$

and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\mathbf{z}t_r \log |x_r|} \operatorname{sgn} x_1 \operatorname{sgn} x_2 dF(x_1, x_2) = s_3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\mathbf{z}t_r \log |x_r|} \operatorname{sgn} x_1 \operatorname{sgn} x_2 dH(x_1, x_2),$$

with  $s_2 = \pm 1$ ,  $s_3 = \pm 1$ .

$$\text{Set} \quad \varphi_1(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\mathbf{z}t_r \log |x_r|} \operatorname{sgn} x_1 dF(x_1, x_2)$$

$$\varphi_2(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\mathbf{z}t_r \log |x_r|} \operatorname{sgn} x_2 dF(x_1, x_2)$$

$$\varphi_{12}(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\mathbf{z}t_r \log |x_r|} \operatorname{sgn} x_1 \operatorname{sgn} x_2 dF(x_1, x_2)$$

and let  $\psi_1(t_1, t_2)$ ,  $\psi_2(t_1, t_2)$ , and  $\psi_{12}(t_1, t_2)$  denote the corresponding transforms of  $H(x_1, x_2)$ . We have

$$(10) \quad \begin{cases} \varphi_1(t_1, t_2) = s_1 \psi_1(t_1, t_2) \\ \varphi_2(t_1, t_2) = s_2 \psi_2(t_1, t_2) \\ \varphi_{12}(t_1, t_2) = s_3 \psi_{12}(t_1, t_2) \end{cases}$$

with  $s_1 = \pm 1$ ,  $s_2 = \pm 1$ , and  $s_3 = \pm 1$ .

Now, as in (9), by considering  $E \left[ e^{i\mathbf{z}t_r \beta (\log |x_r^2| - \log |x_r^2|)} \operatorname{sgn} \left( \frac{x_1^1}{x_1^3} \right) \operatorname{sgn} \left( \frac{x_2^2}{x_2^3} \right) \right]$  we obtain the relation

$$\begin{aligned} \varphi_1(t_{11}, t_{21}) \varphi_2(t_{12}, t_{22}) \varphi_{12}(-t_{11} - t_{12}, -t_{21} - t_{22}) \\ = \psi_1(t_{11}, t_{21}) \psi_2(t_{12}, t_{22}) \psi_{12}(-t_{11} - t_{12}, -t_{21} - t_{22}), \end{aligned}$$

showing that  $s_1, s_2, s_3$ , may be chosen so that  $s_1 s_2 s_3 = 1$ , that is,  $s_1 s_2 = s_3$ .

Consider now the variates  $z'_r = s_r z_r$ ,  $r = 1, 2$ . Let  $K(z'_1, z'_2)$  be the distribution function of  $z'_1, z'_2$ . If we let  $\theta_1(t_1, t_2)$ ,  $\theta_2(t_1, t_2)$ , and  $\theta_{12}(t_1, t_2)$  be the transforms of  $K$  which correspond to  $\varphi_1(t_1, t_2)$ ,  $\varphi_2(t_1, t_2)$ , and  $\varphi_{12}(t_1, t_2)$  respectively, it is evident that

$$(11) \quad \begin{aligned} \varphi_1(t_1, t_2) &= \theta_1(t_1, t_2) \\ \varphi_2(t_1, t_2) &= \theta_2(t_1, t_2) \\ \varphi_{12}(t_1, t_2) &= \theta_{12}(t_1, t_2). \end{aligned}$$

Moreover, from (8),

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i \sum t_r \log |z_r|} dF(x_1, x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i \sum t_r \log |z_r|} dK(x_1, x_2).$$

The last relation, together with the equations (11) imply that  $F(x)$  and  $K(x)$  coincide in each quadrant, thus  $F(x_1, x_2) \equiv K(x_1, x_2)$  for all  $x_1, x_2$ .

The final result is that  $z'_1, z'_2$  have the same distribution as  $x_1, x_2$ , i.e.,  $s_1 e^{a_1} y_1$  and  $s_2 e^{a_2} y_2$  have the same joint distribution as  $x_1$  and  $x_2$ .

The next result bears the same relation to Theorem IIb that Theorem Ic bears to Theorem Ib, that is, only *positive* scale changes are to be permitted.

**THEOREM IIC:** Let  $x_1, \dots, x_k$  have distribution  $F(x_1, \dots, x_k)$  such that the zeros of  $\int e^{i \sum t_r \log |x_r|} dF(x_1, \dots, x_k)$  are nowhere dense, and let  $y_1, \dots, y_k$  have distribution  $G(y_1, \dots, y_k)$ . Let  $(x_1^\alpha, \dots, x_k^\alpha)$  and  $(y_1^\alpha, \dots, y_k^\alpha)$ ,  $\alpha = 1, 2, \dots, n$ , be samples with  $n \geq 3$ . Express  $x_1^\alpha, \dots, x_k^\alpha$  and  $y_1^\alpha, \dots, y_k^\alpha$  in spherical coordinates

$$\begin{aligned} x_i^1 &= r_i \cos \theta_i^1, & y_i^1 &= R_i \cos \varphi_i^1, \\ x_i^2 &= r_i \sin \theta_i^1 \cos \theta_i^2, & y_i^2 &= R_i \sin \varphi_i^1 \cos \varphi_i^2, \\ &\vdots & &\vdots \\ x_i^n &= r_i \sin \theta_i^1 \dots \sin \theta_i^{n-1}; & y_i^n &= R_i \sin \varphi_i^1 \dots \sin \varphi_i^{n-1}. \end{aligned}$$

Then  $\{\theta_i^\beta\}$ ,  $i = 1, \dots, k$ ,  $\beta = 1, \dots, n-1$ , have the same joint distribution as  $\{\varphi_i^\beta\}$  if and only if there exist constants  $k_i > 0$ ,  $i = 1, \dots, k$ , such that the set  $k_i y_i$  have the same joint distribution as the set  $x_i$ .

**PROOF:** If  $\{\theta_i^\beta\}$  have the same distribution as  $\{\varphi_i^\beta\}$  then it follows that  $\left\{ \frac{x_i^\beta}{x_i^n} \right\}$

have the same distribution as  $\left\{ \frac{y_i}{y_i^n} \right\}$ , hence by Theorem IIb there exist constants

$c_i$  such that  $\{c_i y_i\}$  have the same distribution as  $\{x_i\}$ . Set  $z_i = |c_i| y_i$ ; we wish to show that  $\{z_i\}$  have the same distribution as  $\{x_i\}$ . By equation (8) in Theorem IIb it is known that  $\{|z_i|\}$  have the same distribution as  $\{|x_i|\}$ , moreover, if we express  $z_i^\alpha$  in spherical coordinates, the angular coordinates are

the same as those of  $y_i^\alpha$ , therefore  $\left\{\frac{x_i^1}{x_i^2}\right\}$  have the same distribution as  $\left\{\frac{z_i^1}{z_i^2}\right\}$ , since these functions are obtainable in terms of the angular coordinates.

As before, we shall continue the proof from here under the assumption that  $k = 2$ . The procedure is a generalization of the procedure in the proof of Theorem Ic.  $\operatorname{sgn} x_i^1 = \operatorname{sgn} \left\{\frac{x_i^1}{x_i^2}\right\}$ , and similarly for  $y$ , therefore

$$(12) \quad \begin{aligned} & \int \int e^{i \sum_{i=1}^2 t_r (\log |x_i^1| - \log |x_i^2|)} \operatorname{sgn} x_i^1 dF(x_1^1, x_2^1) dF(x_1^2, x_2^2) \\ &= \int \int e^{i \sum_{i=1}^2 t_r (\log |x_i^1| - \log |x_i^2|)} \operatorname{sgn} x_i^1 dH(x_1^1, x_2^1) dH(x_1^2, x_2^2), \quad i = 1, 2, \end{aligned}$$

where it is assumed that  $z_1, z_2$  have distribution  $H(z_1, z_2)$ . As before, set

$$\begin{aligned} \varphi(t_1, t_2) &= \int e^{i \sum_{i=1}^2 t_r \log |x_r|} dF(x_1, x_2), \\ \varphi_i(t_1, t_2) &= \int e^{i \sum_{i=1}^2 t_r \log |x_r|} \operatorname{sgn} x_i dF(x_1, x_2), \quad i = 1, 2, \\ \varphi_{12}(t_1, t_2) &= \int e^{i \sum_{i=1}^2 t_r \log |x_r|} \operatorname{sgn} x_1 \operatorname{sgn} x_2 dF(x_1, x_2), \end{aligned}$$

and denote the corresponding transforms of  $H(x_1, x_2)$  by  $\theta(t_1, t_2)$ ,  $\theta_1(t_1, t_2)$ ,  $\theta_2(t_1, t_2)$ , and  $\theta_{12}(t_1, t_2)$ . It has been remarked already that  $\{|z_i|\}$  have the same distribution as  $\{|x_i|\}$ , therefore  $\theta(t_1, t_2) = \varphi(t_1, t_2)$ . Equation (12) yields the relation  $\varphi_i(t_1, t_2)\varphi(-t_1, -t_2) = \theta_i(t_1, t_2)\theta(-t_1, -t_2)$ ,  $i = 1, 2$ ; the zeros of  $\varphi(t_1, t_2)$  are nowhere dense, so that it can be concluded that  $\varphi_i(t_1, t_2) = \theta_i(t_1, t_2)$ ,  $i = 1, 2$ . Now, from an equation similar to (12) we obtain  $\varphi_{12}(t_1, t_2) = \theta_{12}(t_1, t_2)$ . As in Theorem IIb, the four relations above together imply that  $F(x_1, x_2) \equiv H(x_1, x_2)$ , in other words,  $\{|c_i|y_i\}$  have the same distribution as  $\{x_i\}$ .

We are now in a position to combine some of the preceding theorems so as to obtain analogous results for scale and location parameters together.

**THEOREM IIIA:** *Let  $x$  have distribution  $F(x)$  such that the zeros of  $\int e^{itx} dF(x)$  satisfy the condition of Theorem Ia, and the zeros of*

$$\int \int \int e^{it_1 \log |x_1 - x_3| + it_2 \log |x_2 - x_3|} dF(x_1) dF(x_2) dF(x_3)$$

*are nowhere dense, and let  $y$  have distribution  $G(y)$ . Let  $x_1, \dots, x_n$  and  $y_1, \dots, y_n$  be samples, with  $n \geq 9$ . Then  $w_\alpha = \frac{x_\alpha - x_n}{x_{n-1} - x_n}$ ,  $\alpha = 1, \dots, n-2$ ,*

*have the same joint distribution as the corresponding set  $w'_\alpha = \frac{y_\alpha - y_n}{y_{n-1} - y_n}$  if and only if there exist constants  $a, c$ , such that  $c(y - a)$  and  $x$  have the same distribution.*

PROOF: Sufficiency of the condition is an immediate consequence of the fact that  $w'_\alpha$  is invariant under transformations of the form  $y' = c(y - a)$ . Assume then that  $\{w_\alpha\}$  and  $\{w'_\alpha\}$  have the same joint distribution. By elementary transformations it is evident that the functions  $\frac{x_1 - x_3}{x_7 - x_9}, \frac{x_4 - x_6}{x_7 - x_9}, \frac{x_2 - x_3}{x_8 - x_9}, \frac{x_5 - x_4}{x_8 - x_9}$ , have the same joint distribution as the corresponding functions of the  $y$ 's, if  $n \geq 9$ . Since  $x_1, \dots, x_n$  form a sample it follows that the pairs  $\{x_1 - x_3, x_2 - x_3\}$ ,  $\{x_4 - x_6, x_5 - x_6\}$ ,  $\{x_7 - x_9, x_8 - x_9\}$ , have the same joint distributions and are pairwise independent, and similarly for the corresponding functions of the  $y$ 's. Theorem IIb assures the existence of constants  $c_1, c_2$ , such that  $c_1(y_1 - y_3), c_2(y_2 - y_3)$  have the same joint distribution as  $(x_1 - x_3), (x_2 - x_3)$ . Considering separately the marginal distributions it is seen that  $c_1(y_1 - y_3)$  has the same distribution as  $c_2(y_2 - y_3)$ .  $y_1 - y_3$  and  $y_2 - y_3$  have the same distribution, therefore either  $c_2 = c_1$ , or  $c_2 = -c_1$ . Set  $u_\alpha = x_\alpha - x_3$ ,  $v_\alpha = c_1(y_\alpha - y_3)$ ,  $\alpha = 1, 2$ . We have, for the distributions of  $(u_1, u_2)$  and  $(v_1, v_2)$ , relations corresponding to (10) in Theorem IIb, with the additional condition that  $s_1 = s_2$ , because of the symmetry in the variables. This implies that either  $(v_1, v_2)$  or  $(-v_1, -v_2)$  have the same joint distribution as  $(u_1, u_2)$ , that is, there exists  $c$  such that  $c(y_1 - y_3)$  and  $c(y_2 - y_3)$  have the same joint distribution as  $x_1 - x_3$  and  $x_2 - x_3$ . Application of Theorem Ia now completes the proof.

Just as before, there is an analogous situation when we consider angular coordinates instead of quotients. The proof is immediate; the angular coordinates determine the angular coordinates of  $\{x_1 - x_3, x_2 - x_3\}$ ,  $\{x_4 - x_6, x_5 - x_6\}$ , and  $\{x_7 - x_9, x_8 - x_9\}$ , arranged as a sample. Then the constants  $c_1, c_2$  in the proof of Theorem IIIa are both positive; it follows that  $c_1 = c_2$ . Application of Theorem Ia gives

THEOREM IIIB: Let  $x_1, \dots, x_n$  and  $y_1, \dots, y_n$  satisfy the hypotheses of Theorem IIIa. Set

$$\begin{aligned} x_1 - x_n &= r \cos \theta_1, & y_1 - y_n &= r' \cos \theta'_1, \\ x_2 - x_n &= r \sin \theta_1 \cos \theta_2, & y_2 - y_n &= r' \sin \theta'_1 \cos \theta'_2, \end{aligned}$$

$$x_{n-1} - x_n = r \sin \theta_1 \dots \sin \theta_{n-2}; \quad y_{n-1} - y_n = r' \sin \theta'_1 \dots \sin \theta'_{n-2}.$$

Then  $\theta_1, \dots, \theta_{n-2}$  have the same joint distribution as  $\theta'_1, \dots, \theta'_{n-2}$  if and only if there exist constants  $a$  and  $c > 0$  such that  $c(y - a)$  has the same distribution as  $x$ .

Theorem IVa is a generalization of Theorem Ia to cover arbitrary linear combinations of some subset of the sample.

THEOREM IVa: Suppose  $x$  has distribution  $F(x)$  such that  $\int e^{itx} dF(x)$  does not vanish, and let  $y$  have distribution  $G(y)$ . Consider the functions  $w_\alpha = x_\alpha - \sum_{\beta=1}^{n-m} l_{\alpha\beta} x_{m+\beta}$ ,  $w'_\alpha = y_\alpha - \sum_{\beta=1}^{n-m} l_{\alpha\beta} y_{m+\beta}$ ,  $\alpha = 1, 2, \dots, m$ ,  $\beta = 1, 2, \dots$ ,



$n - m$ , and suppose that  $m > n - m$ . Then, if  $\{w_\alpha\}$  have the same joint distribution as  $\{w'_\alpha\}$  and if  $\sum_{\beta=1}^{n-m} l_{\alpha\beta} \neq 1$  for some  $\alpha$ , it follows that  $F(y) \equiv G(y)$ ; if  $\sum_{\beta} l_{\alpha\beta} = 1$  for all  $\alpha$  there exists a constant  $a$  such that  $F(y - a) \equiv G(y)$ .

PROOF: Denote the characteristic functions of  $x$  and  $y$  by  $\varphi(t)$  and  $\psi(t)$  respectively. By expressing the fact that  $\{w_\alpha\}$  and  $\{w'_\alpha\}$ ,  $\alpha = 1, 2, \dots, n - m + 1$ , have the same characteristic function we obtain the functional equation

$$\prod_{\alpha=1}^{n-m+1} \varphi(t_\alpha) \prod_{\beta=1}^{n-m} \varphi\left(-\sum_{\alpha=1}^{n-m+1} l_{\alpha\beta} t_\alpha\right) = \prod_{\alpha=1}^{n-m+1} \psi(t_\alpha) \prod_{\beta=1}^{n-m} \psi\left(-\sum_{\alpha=1}^{n-m+1} l_{\alpha\beta} t_\alpha\right).$$

By hypothesis  $\varphi(t)$  does not vanish, therefore  $\psi(t)$  has no zeros, because of the relation above.  $\varphi(t)$  and  $\psi(t)$  are continuous, thus the function  $f(t) = \log \varphi(t) - \log \psi(t)$  can be uniquely defined in a continuous manner for all  $t$ . The equation above becomes

$$(13) \quad \sum_{\alpha=1}^{n-m+1} f(t_\alpha) + \sum_{\beta=1}^{n-m} f\left(-\sum_{\alpha=1}^{n-m+1} l_{\alpha\beta} t_\alpha\right) = 0.$$

The constants  $l_{\alpha\beta}$  are necessarily linearly dependent, so that, for some  $\alpha$ ,  $l_{\alpha\beta}$  can be expressed as a linear combination of the others; suppose then that

$$l_{n-m+1,\beta} = \sum_{\alpha=1}^{n-m} e_\alpha l_{\alpha\beta}.$$

Putting these values in (13) we have

$$(14) \quad \sum_{\alpha=1}^{n-m+1} f(t_\alpha) + \sum_{\beta=1}^{n-m} f\left(-\sum_{\alpha=1}^{n-m} l_{\alpha\beta}(t_\alpha + t_{n-m+1}e_\alpha)\right) = 0.$$

It can be assumed that  $\sum e_\alpha^2 \neq 0$ , for, if  $e_\alpha = 0$  for all  $\alpha$ , we have  $l_{n-m+1,\beta} = 0$ ,  $\beta = 1, \dots, n - m$ , that is,  $w'_{n-m+1} = y_{n-m+1}$  and  $w_{n-m+1} = x_{n-m+1}$ , hence  $x$  and  $y$  have the same distribution. Assuming  $e_1 \neq 0$ , set  $t_\alpha = -e_\alpha t_{n-m+1}$ ,  $\alpha = 2, \dots, n - m$ , in (14), obtaining

$$(15) \quad f(t_1) + \sum_{\alpha=2}^{n-m} f(-e_\alpha t_{n-m+1}) + f(t_{n-m+1}) + \sum_{\beta=1}^{n-m} f(-l_{1\beta}(t_1 + e_1 t_{n-m+1})) = 0,$$

now, recalling that  $f(0) = 0$ , set  $t_{n-m+1} = 0$ , getting  $f(t_1) + \sum_{\beta=1}^{n-m} f(-l_{1\beta}t_1)$ .

Evaluating this with argument  $t_1 + e_1 t_{n-m+1}$ , and substituting back in (15) it appears that

$$(16) \quad f(t_1) + f(t_{n-m+1}) + \sum_{\alpha=2}^{n-m} f(-e_\alpha t_{n-m+1}) = f(t_1 + e_1 t_{n-m+1}).$$

Now setting  $t_1 = 0$  in (16) we have the relation

$$f(t_{n-m+1}) + \sum_{\alpha=2}^{n-m} f(-e_\alpha t_{n-m+1}) = f(e_1 t_{n-m+1}).$$

thus we have finally  $f(t_1) + f(e_1 t_{n-m+1}) = f(t_1 + e_1 t_{n-m+1})$ , or, since  $e_1 \neq 0$ ,  $f(t_1 + t_2) = f(t_1) + f(t_2)$ . The last relation implies that  $f(t) = ct$ , since  $f(t)$  is continuous. Now replace  $f(t)$  by  $ct$  in (13), getting  $c \left\{ \sum_{\alpha=1}^{n-m+1} t_\alpha - \sum_{\alpha=1}^{n-m+1} \sum_{\beta=1}^{n-m} l_{\alpha\beta} t_\alpha \right\} = 0$ , that is, either  $c = 0$ , or  $\sum_{\beta=1}^{n-m} l_{\alpha\beta} = 1$  for all  $\alpha$ . We conclude then that  $\varphi(t) = \psi(t)$ , unless  $\sum_{\beta} l_{\alpha\beta} = 1$  for all  $\alpha$ . If  $\sum_{\beta} l_{\alpha\beta} = 1$  for all  $\alpha$  we have  $\varphi(t) = e^{at} \psi(t)$ .  $\varphi(-t) = \varphi(t)$  and  $\psi(-t) = \psi(t)$ , hence  $c$  is of the form  $c = ia$ , where  $a$  is real, in other words  $\varphi(t) = e^{iat} \psi(t)$ , thus concluding the proof of the theorem.

It was assumed in Theorem IVa that  $\varphi(t)$  has no zeros. If  $\varphi(t)$  has zeros we have proved that, for an interval  $|t| < \epsilon$ ,  $\varphi(t) = \psi(t)$  (or  $\varphi(t) = e^{iat} \psi(t)$ ). This does not necessarily imply the result of Theorem IVa, but it does imply at least that if the  $k$ th moments of  $x$  and of  $y$  (or of  $y - a$ ) both exist they are equal.

The last result in this series can be proved by methods similar to those used in Theorem IVa.

**THEOREM IVB:** *Let  $x$  and  $y$  satisfy the hypotheses of Theorem IVa. Suppose, moreover, that  $m > 2(n - m)$ , that the rank of  $\|l_{\alpha\beta}\|$  is  $n - m$ , and that  $\sum_{\beta=1}^{n-m} l_{\alpha\beta} \neq 1$  for at least  $2m - n$  values of  $\alpha$ . Then, if there exist constants  $\{c_\alpha\}$  such that the set  $\{c_\alpha w'_\alpha\}$  have the same joint distribution as  $\{w_\alpha\}$ , it follows that, for some  $\alpha$ ,  $c_\alpha y$  has the same distribution as  $x$ .*

**3. Application to Composite Hypotheses.** The results of section 2 have a significant application in the theory of testing composite hypotheses. Suppose that  $x$  has a distribution of the form  $F(x, \theta_1, \theta_2)$ , and that the hypothesis  $\theta_2 = \theta_2^0$  is to be tested, without reference to the value of  $\theta_1$ . We assume that the parameters are independent, i.e.,  $F(x, \theta_1, \theta_2) \equiv F(x, \theta'_1, \theta'_2)$  implies that  $\theta_1 = \theta'_1$  and  $\theta_2 = \theta'_2$ . It is true in a wide class of important cases that, given a sample  $x_1, \dots, x_n$  from the distribution  $F(x, \theta_1, \theta_2)$ , there exist functions  $y_\alpha(x_1, \dots, x_n)$ ,  $\alpha = 1, 2, \dots, p$ , such that  $\{y_\alpha\}$  have joint distribution independent of  $\theta_1$ , but depending on  $\theta_2$ . Now if the  $\{y_\alpha\}$  are such that their joint distribution redetermines the original distribution, except for  $\theta_1$ , one can reasonably use the  $p$ -dimensional distribution of the  $\{y_\alpha\}$  for testing the hypothesis  $\theta_2 = \theta_2^0$ , thus reducing the composite hypothesis to a simple hypothesis. In testing this simple hypothesis, every alternative hypothesis (corresponding to a value of  $\theta_2$ ) determines a distribution of  $x$  among the alternatives  $F(x, \theta_1, \theta_2)$  except for the unknown  $\theta_1$ , that is, there is a one-to-one correspondence between the two sets of alternative hypotheses, expressed by the fact that if  $\theta'_2 = \theta_2''$  then the distributions of the set  $\{y_\alpha\}$  corresponding to  $\theta_2 = \theta'_2$  and  $\theta_2 = \theta_2''$  must be different.

Suppose, for example, that it is desired to test whether  $y = x - a$  for some  $a$  has the distribution  $F(y, \theta^0)$ , with the assumption that, for some  $a$ ,  $y$  has the

distribution  $F(y, \theta)$ . Given a sample one can form the set  $w_\alpha = x_\alpha - x_n$ ,  $\alpha = 1, 2, \dots, n-1$ , obtaining the distribution  $G(w_1, \dots, w_{n-1}, \theta)$ ; now consider the simple hypothesis  $\theta = \theta^0$ , knowing that  $G$  determines  $\theta$ , by Theorem Ia. Similarly one can test whether  $cx$ , for some  $c \neq 0$ , has distribution  $F(y, \theta^0)$ , by forming  $w_\alpha = x_\alpha/x_n$ ,  $\alpha = 1, \dots, n-1$ , or by expressing  $(x_1, \dots, x_n)$  in spherical coordinates and considering the angular coordinates, according to whether both positive and negative or only positive values of  $c$  are to be allowed.

In the same way one can test the hypothesis  $\theta = \theta^0$  under the assumption that  $c(x-a)$  has distribution  $F(y, \theta)$  by forming  $w_\alpha = \frac{x_\alpha - x_n}{x_{n-1} - x_n}$ ,  $\alpha = 1, \dots, n-2$ , or by expressing  $(x_1 - x_n, \dots, x_{n-1} - x_n)$  in spherical coordinates and considering the angular coordinates.

Theorem IVa may be applied to analogous problems, in which the hypothesis  $\theta = \theta^0$  is to be tested under the assumption that  $y = u - \sum a_i x_i$  has distribution  $F(y, \theta)$  for fixed values of the  $x_i$ , with the  $a_i$  unknown. In such problems there exist linear combinations of the observed values of  $y$  which are independent of the  $a_i$ . By Theorem IVa, under certain conditions the joint distribution of these linear combinations determines the original distribution of  $y$ , without regard to the  $a_i$ .

In applying some of the preceding results we must verify in certain cases that the zeros of  $\int e^{itz} dF(x)$  are nowhere dense, for a certain distribution function.

By a change of variable the condition of Theorem Ib can be stated in this form; moreover if  $F(x)$  satisfies this condition it is evident that it satisfies the condition of Theorem Ia. A sufficient condition applicable to a considerable class of cases has been obtained by Levinson [4]; if  $f(x)$  is  $O(e^{-\theta(x)})$  as  $x \rightarrow \infty$ , where  $\theta(x)$  is monotone and  $\int_1^\infty \frac{\theta(x)}{x^2} dx$  diverges to  $\infty$ , then  $\int e^{itz} f(x) dx$  cannot vanish on an interval without vanishing identically. It is evident that it is likewise sufficient if the corresponding condition holds as  $x \rightarrow -\infty$  instead of  $+\infty$ . In particular, if there exists  $A$  such that  $f(x) = 0$  for  $x > A$  (or for  $x < A$ ) it is a consequence of the Levinson result that  $\int e^{itz} f(x) dx$  has no intervals of zeros. It can be established easily that if  $f(x)$  is majorized by  $|x|^{-(1-\epsilon)}$ ,  $\epsilon > 0$ , in the neighborhood of the origin, then  $\int e^{it \log |x|} f(x) dx$  has no intervals of zeros.

As a simple example consider the rectangular distribution on  $(0, 1)$ . Let  $(x-a)/r$  have this distribution with  $a$  unknown,  $r > 0$ , and suppose that we are interested only in  $r$ . Given a sample  $x_1, \dots, x_n$  form the functions  $y_\alpha = (x_\alpha - x_n)/r$ ,  $\alpha = 1, \dots, n-1$ . Set  $y_M = \max(y_\alpha, 0)$ ,  $y_L = \min(y_\alpha, 0)$ . Then it can be shown that  $y_1, \dots, y_{n-1}$  have probability density  $(1 - y_M + y_L)$  in the region  $-1 \leq y_\alpha \leq 1$ ,  $y_M - y_L \leq 1$ , zero elsewhere.  $\psi = y_M - y_L$  is of course the quotient of the sample range by  $r$ . It can be shown that  $\psi$  has

density  $n(n-1)(1-\psi)\psi^{n-2}d\psi$ . Theorem Ia makes it possible to base any tests not involving  $a$  on the distribution of the  $y_\alpha$ , since if the  $y_\alpha$  have the stated distribution then  $(x-a)/r$  for some  $a$  must have the rectangular distribution.

Similarly, suppose  $y = (x-a)/r$  has the distribution  $e^{-y}$ ,  $y > 0$ , for some  $a, r$ . Then  $w_\alpha = \frac{x_\alpha - x_n}{r}$ ,  $\alpha = 1, 2, \dots, n-1$ , have distribution density  $\frac{1}{n} e^{-\sum w_\alpha + nw_L}$ , where  $w_L = \min(0, w_\alpha)$ . Again, the latter distribution may be used to estimate  $r$ .

Let us examine the distributions of functions of the type considered, in the case of normality. Assume that  $x_1, \dots, x_n$  are a sample of  $n$  observations from a normal distribution with unit variance and unknown mean. The variables  $y_\alpha = x_\alpha - x_1$ ,  $\alpha = 2, \dots, n$ , have a joint normal distribution with zero means and matrix of variances and covariances  $\|A^{ij}\| = \|1 + \delta_{ij}\|$ . Then Theorem Ia shows that if  $\{y_\alpha\}$  have this joint distribution then  $x$  is normally distributed with unit variance. Note that  $\chi_{n-1}^2 \equiv \sum A_{ij} y_i y_j \equiv \sum (x_\alpha - \bar{x})^2$ . If we had  $x = x'/\sigma$ , then  $\sum (x'_\alpha - \bar{x}')^2 = \sigma^2 \chi_{n-1}^2$ , giving the estimate  $\frac{1}{n-1} \sum (x'_\alpha - \bar{x}')^2$  for  $\sigma^2$ .

There are, of course, many ways in which the matrix  $\|A_{ij}\|$  may be transformed into a diagonal matrix in order to obtain a new set of independently distributed variates; one convenient set is the set  $\sqrt{\frac{1}{2}} y_2, \sqrt{\frac{2}{3}} (y_3 - \frac{1}{2} y_2), \dots, \sqrt{\frac{n-1}{n}} \left( y_n - \frac{1}{n-1} \sum_{\alpha=2}^{n-1} y_\alpha \right)$ . In terms of the original  $x$ 's we have  $\sqrt{\frac{1}{2}} (x_2 - x_1), \sqrt{\frac{2}{3}} (x_3 - \frac{1}{2}(x_1 + x_2)), \sqrt{\frac{n-1}{n}} \left( x_n - \frac{1}{n-1} \sum_{\alpha=1}^{n-1} x_\alpha \right)$ ; these functions of the data are independently distributed according to the normal distribution with zero mean and unit variance.

Similarly, in the case of a sample  $x_1, \dots, x_n$  from a normal distribution with zero mean and unknown variance, there exists a set of  $n-1$  functions with distributions independent of the variance. A convenient set of functions is the set

$$t_m = \frac{\sqrt{m} x_{m+1}}{\sqrt{\sum_{i=1}^m x_i^2}}; \quad m = 1, \dots, n-1.$$

It is known (see Bartlett [1]) that the variables  $t_m$  are independently distributed according to student  $t$ -distributions with  $m$  degrees of freedom respectively. The set  $t_m$  determines the set of angular coordinates obtained by expressing  $x_1, \dots, x_n$  in spherical coordinates, hence we can conclude, conversely, that if  $\{t_m\}$  have this joint distribution then  $x$  is normal with mean zero.

Finally we can eliminate both mean and variance. Suppose  $x_1, \dots, x_n$  are a sample from some normal distribution. The variables

$$u_m = \sqrt{\frac{m}{m+1}} \left\{ x_{m+1} - \frac{1}{m} \sum_{i=1}^m x_i \right\}, \quad m = 1, 2, \dots, n-1,$$

are normal and independent with mean zero and some variance. Then we have the set

$$t'_r = \frac{\sqrt{r \left( \frac{r+1}{r+2} \right)} \left\{ x_{r+2} - \frac{1}{r+1} \sum_{i=1}^{r+1} x_i \right\}}{\sqrt{\sum_{j=1}^r \frac{j}{j+1} \left\{ x_{j+1} - \frac{1}{j} \sum_{i=1}^j x_i \right\}^2}}, \quad r = 1, \dots, n-2,$$

independently distributed according to  $t$ -distributions with  $r$  degrees of freedom respectively. It may be convenient for computational purposes to make use of the identity

$$\sum_{j=1}^r \frac{j}{j+1} \left\{ x_{j+1} - \frac{1}{j} \sum_{i=1}^j x_i \right\}^2 \equiv \sum_{j=1}^{r+1} \left( x_j - \frac{1}{r+1} \sum_{i=1}^{r+1} x_i \right)^2 \equiv \sum_{j=1}^{r+1} (x_j - \bar{x}_{(r+1)})^2.$$

We then have

$$t'_r = \frac{\sqrt{r \left( \frac{r+1}{r+2} \right)} (x_{r+2} - \bar{x}_{(r+1)})}{\sqrt{\sum_{i=1}^{r+1} (x_i - \bar{x}_{(r+1)})^2}}, \quad r = 1, \dots, n-2.$$

Now, by Theorem IIIc, we know that if the set  $\{t'_r\}$  has this specified distribution then  $x$  must be distributed according to some normal distribution. The set  $\{t'_r\}$  may be used to test the goodness of fit of the observations to normality, by first adjusting the set  $\{t'_r\}$  to a standard basis of comparison, i.e., by considering  $F_r(t'_r)$ , where  $F_r$  is the corresponding cumulative distribution function and then applying, for example, a  $\chi^2$  goodness of fit test to these  $n-2$  quantities, with respect to the rectangular distribution on  $(0, 1)$ .

#### REFERENCES

- [1] M. S. BARTLETT, *Proc. Camb. Phil. Soc.* Vol. 30 (1934), pp. 164-169.
- [2] J. NEYMAN AND E. S. PEARSON, *Biometrika*, Vol. XXA, (1928) pp. 175-240, 264-294.
- [3] R. A. FISHER, *Proc. Roy. Soc. A*, Vol. 144 (1934), pp. 285-307.
- [4] N. LEVINSON, *Proc. London Math. Soc.* Vol. 41 (1936), pp. 393-407.
- [5] E. J. G. PITMAN, *Biometrika* Vol. 30 (1939), pp. 391-421.

PRINCETON UNIVERSITY,  
PRINCETON, N. J.

# THE SELECTION OF VARIATES FOR USE IN PREDICTION WITH SOME COMMENTS ON THE GENERAL PROBLEM OF NUISANCE PARAMETERS

BY HAROLD HOTELLING

1. **Maximum Correlation as a Test.** For predicting or estimating a particular variate  $y$  there is frequently available an embarrassingly large number of other variates having some correlation with  $y$ . For example, in fitting demand functions by means of economic time series, the number of series of observations having some relation to the demand which is sought to be estimated is apt to be very large, whereas the number of good independent observations on each is quite small. The proper coefficients in the regression equation must ordinarily be determined from the observations, and must not exceed in number the observations on each variate. Furthermore, in order to have a measure of error that will make it possible to distinguish real effects from those due to chance, it is necessary that the number of predictors<sup>1</sup> shall be enough less than the number of observations on each variate so that the residual chance variance can be determined with an appropriate degree of accuracy. It is desirable to select a set of predictors yielding estimates of maximum but determinable accuracy, and at the same time to avoid the fallacies of selection among numerous results of that one which appears most significant and treating it as if it were the only one examined.

Considerations other than maximum and determinate accuracy are of practical importance. The labor of calculation by the method of least squares becomes a serious obstacle to the use of the theoretically optimum set of variates when these are very numerous, though the rapid current development of mechanical and electrical devices suitable for these computations offers a hope that the limits now set in practice in this way will soon be considerably increased. Furthermore, predictions or estimates must, as in speculative business or in military activity, be made from moment to moment, often in a rough manner by persons incapable of or averse to using complex formulae, and in such activities frequent revisions of the regression equations must be made to accord with altered conditions. Also, in temporal predictions, the time of availability of

<sup>1</sup> I use this term for what are often called the independent variates in a regression equation, since these ordinarily are not really independent in the probability sense. Similarly I shall call the "dependent" variate the *predictand*. By *prediction* I mean merely the use of regression equations to estimate some unknown variate by means of the values of related variates, without any necessary connotation of temporal order, though the most interesting applications seem for the most part to be those in which we pass from a knowledge of the past to an estimate of the future.

the values of the predictors is important, since an early prediction (e.g. of the size of a harvest) is more valuable than a later one of the same accuracy.

If we make the usual assumption<sup>2</sup> that the probability distribution of  $y$  is, for every set of values of the predictors, normal with a fixed variance  $\sigma^2$  and an expectation that is a linear function of the predictors, we shall wish to minimize  $\sigma^2$  subject to appropriate limitations, and this amounts to the same thing as maximizing the multiple correlation  $\rho$  of  $y$  with the predictors, since  $1 - \rho^2$  is the ratio of  $\sigma^2$  to the total variance of  $y$ , which is the same for all sets of predictors. The estimates  $s$  and  $R$  of  $\sigma$  and  $\rho$  obtained from the available sample are of course a different matter. But it is clear that the value of  $R$  provides a suitable criterion of choice under the following conditions: We are called upon to choose one among two or more sets, each consisting of a fixed number of predictors; for each predictor we have a known value corresponding to each of the values  $y_1, \dots, y_N$  observed for the predictand; and there is no basis for preferring one of these sets to another either in theory, in observations extraneous to those just specified, or in cost or time of availability. In particular, if just one predictor is to be used, that having the highest sample correlation with the predictand should under these conditions be the one adopted. But in making such a choice a test of its accuracy is required, to take account of the possibility that the wrong choice has been made because of chance fluctuations in the sample correlation coefficients.

There are innumerable economic variates available for prediction of business conditions, and most of these are highly correlated with each other. The selection of one business index instead of another for a particular purpose will involve the question which has exhibited the higher correlation with the quantity to be predicted, and consequently the question of the definiteness with which the difference between the calculated correlations can be regarded as significant.

Our problem evidently has a bearing on governmental policy in selecting among the numerous series of data those whose continuation will be most valuable. The high cost of assembling these statistics dictates a careful selection of a limited number of series having little correlation with each others' current values, but with correlations as great as possible with those things whose prediction or estimation is most important.

**2. The Choice of one Predictor with Two Available.** Let us take first the simplest case, which may be illustrated by a Michigan State College problem of

<sup>2</sup> We shall not here go into the question of the applicability of these standard assumptions to time series otherwise than to note that some transformations of observations ordered in time are usually necessary and sufficient to obtain quantities satisfying the assumptions so closely that deviations from them cannot be detected. Such transformations include replacing a variate by its logarithm, and eliminating trend and seasonal variations by least squares. In view of the satisfactory adjusted observations found empirically by these and similar methods, the usual objections to studying time series by exact methods seem much exaggerated.

which Dr. W. D. Baten has told me. The ultimate weight of a mature ox is estimated by means of his length at an early age. The question has been raised, however, whether a more accurate prediction might not be made by means of the calf's girth at his heart. Records were at hand of 13 oxen showing their lengths and girths as calves and also their weights when mature. A regression equation involving both length and girth would presumably give greater accuracy than either variate alone; but it appears that those who make the estimates desire a simple formula involving only one variate. Suppose, then, that in such a sample the correlation of weight with length is  $r_1 = .7$ , that the correlation of weight with girth is  $r_2 = .5$ , and that the correlation of girth with length is  $r_0 = .4$ . Is the difference  $r_1 - r_2 = .2$  sufficiently great in relation to its sampling errors to warrant the inference that girth is really a better predictor than length, or must the question be left in abeyance until more observations can be accumulated?

A straightforward procedure which would have been used with little question before the advent of modern exact methods is to calculate the asymptotic approximation to the standard error of  $r_1 - r_2$  by the differential method, assuming the three variates to have the trivariate normal distribution, and to regard the difference of the correlations as significant if it exceeds a multiple of this standard error determined by the tables of the normal distribution. The calculation of the asymptotic approximation  $\sigma_{r_1-r_2}$  may be carried out in the following manner. Let  $\rho_1$ ,  $\rho_2$ , and  $\rho_0$  be the population values of  $r_1$ ,  $r_2$ , and  $r_0$  respectively. Then if  $\sigma_{ij}$  denote the population covariance of  $x_i$  and  $x_j$  ( $i, j = 0, 1, 2$ ), we have

$$\rho_1 =$$

with similar formulae for  $\rho_2$  and  $\rho_0$ . Likewise the sample estimates of these parameters are given by such expressions as

$$r_1 = \frac{s_{01}}{s_{00}}$$

Taking the logarithm of this last expression, expanding about the population values, denoting by the operator  $\delta$  the deviation of sample from population values of the covariances, and the resultant deviation in  $r_1$ , and dropping terms of order higher than the first, we have:

$$\delta r_1 = \rho_1 \left( \frac{\delta s_{01}}{s_{01}} - \frac{\delta s_{00}}{2s_{00}} - \frac{\delta s_{11}}{2s_{11}} \right).$$

In the same way

$$\delta r_2 = \rho_2 \left( \frac{\delta s_{02}}{s_{02}} - \frac{\delta s_{00}}{2s_{00}} - \frac{\delta s_{22}}{2s_{22}} \right).$$

The asymptotic value of the sampling covariance is obtained by multiplying these two expressions together and taking the expectation. The sampling covariance of two estimates of covariance of the usual kind (sum of products



divided by number of degrees of freedom) in the same sample, having  $n$  degrees of freedom (which ordinarily means that there are  $n + 1$  individuals in the sample and that the means are eliminated), is given exactly by the formula<sup>3</sup>

$$E(\delta s_{ij} \delta s_{km}) = (\sigma_{ik} \sigma_{jm} + \sigma_{im} \sigma_{jk})/n,$$

in which the subscripts may have any values, equal or unequal. When this formula is applied to each of the nine terms of the product and the results are expressed in terms of the correlations  $\rho_i$ , there results the asymptotic expression for the covariance given by

$$nE(\delta r_1 \delta r_2) = \frac{1}{2} \rho_1 \rho_2 (\rho_1^2 + \rho_2^2 + \rho_0^2 - 1) + \rho_0 (1 - \rho_1^2 - \rho_2^2).$$

This method provides also one of the derivations of the familiar formula which may be written

$$n\sigma_{r_1}^2 = nE(\delta r_1)^2 = (1 - \rho_1^2)^2, \quad n\sigma_{r_2}^2 = (1 - \rho_2^2)^2.$$

The variance of the difference of  $r_1$  and  $r_2$  is the sum of their variances minus twice their covariance. Hence

$$n\sigma_{r_1-r_2}^2 = (1 - \rho_1^2)^2 + (1 - \rho_2^2)^2 - \rho_1 \rho_2 (\rho_1^2 + \rho_2^2 + \rho_0^2 - 1) + 2\rho_0 (\rho_1^2 + \rho_2^2 - 1).$$

We are testing the hypothesis that  $\rho_1 = \rho_2$ . If we put a common value  $\rho$  for them in the last expression and simplify, we obtain for the standard error of the difference,

$$\sigma_{r_1-r_2} = \sqrt{\frac{(1 - \rho_0)(2 - 3\rho^2 + \rho_0\rho^2)}{n}}.$$

The second factor in parentheses is always positive because of the inequalities limiting the correlations among three variates.

This formula contains two unknown parameters,  $\rho$  and  $\rho_0$ . The classical procedure would be substitute  $r_1$ ,  $r_2$  and  $r_0$  respectively for  $\rho_1$ ,  $\rho_2$ , and  $\rho_0$  in the previous formula, and use the resulting standard error expression as if the ratio to it of  $r_1 - r_2$  were normally distributed. A first modification, more in line with modern ideas, would be to use some kind of average of  $r_1$  and  $r_2$  as an estimate of both  $\rho_1$  and  $\rho_2$ , since the null hypothesis tested is that these are equal. But whatever sample estimates we substitute for  $\rho$  and  $\rho_0$ , the formula remains unsatisfactory, since no suitable limits of error are available. If instead of the standard error we were to work out the exact distribution of  $r_1 - r_2$  we should still not be free from the difficulty. This exact distribution clearly involves both  $\rho$  and  $\rho_0$ , since its variance does so. Neither can we escape from the trouble by using some function  $z = f(r)$ , such as the inverse hyperbolic tangent suggested by R. A. Fisher, and considering the standard error of  $z_1 - z_2 =$

<sup>3</sup> I have given a derivation of this formula from the characteristic function of the multivariate normal distribution [1]. Numerous special cases appear in earlier literature. The derivation above is a simplification and improvement of several versions, appearing in the various early writings of Karl Pearson.

$f(r_1) - f(r_2)$ ; for this standard error will have as the first term in its expansion in a series of powers of  $n^{-1}$  simply the product of the expression above for  $\sigma_{r_1-r_2}$  by  $f'(\rho)$ ; and this must clearly involve both  $\rho_0$  and  $\rho$ .

**3. Nuisance Parameters.** This is not by any means the only statistical problem in which unknown and undesired parameters enter into the distribution of the statistic which we should naturally use to test a hypothesis. Indeed, the early investigation which was perhaps most influential in setting the whole tone of modern statistical research was that [2] in which W. C. Gosset ("Student") arrived at the exact distribution of the ratio of a deviation in the mean to the *estimated* standard error. The previous practice (which unfortunately survives today in some quarters, and is even taught to students without explaining its approximate character) was to neglect the sampling errors in the estimate of the unknown variance  $\sigma^2$  and to treat the ratio as normally distributed with unit variance. The rigorous derivation by Fisher [3] of the Student distribution makes clear the manner in which the nuisance parameter  $\vartheta$  may in this, and in some other, problems be eradicated from the distribution through integration, after altering the original statistic (the deviation in the mean) by dividing it by another statistic. The new statistic, the Student ratio, vanishes whenever the old statistic, the deviation in the mean, does so, and the same hypothesis is tested by both. This then is one way to get rid of a nuisance parameter: when you have a statistic estimating a parameter whose vanishing is in question, but whose distribution involves another parameter, alter the statistic by multiplying or dividing by another statistic in such a way that the new function vanishes whenever the old one does so; and *do this in such a way that the new distribution will be independent of the nuisance parameter*. Unhappily, this method has been applied successfully only in particular cases, and no way to use it in the problem at hand has been found.

A second method is that of transformation employed by Fisher in dealing with such problems as testing the significance of the difference between the correlation coefficients in independent samples between the same two variates. The need for the transformation in this case is occasioned by the presence in the distribution of the difference of the sample correlations of the unknown true value, which is not directly relevant to the comparison. We have seen that this method also fails to solve our problem.

A third method of dealing with nuisance parameters is the use of fiducial probability by R. A. Fisher [4] and by Daisy M. Starkey [5] in testing the significance of the difference between the means of two samples when the variances may be unequal. Criticisms of these applications of fiducial probability have been made by M. S. Bartlett [6] and B. L. Welch [7], and the field of applicability of such methods is still in need of elucidation.

Some findings of J. Neyman [8] having a bearing on the general nuisance parameter problem should also be noted.

The only other class of methods for dealing with nuisance parameters of which

I am aware involves the comparison of the particular sample obtained, not with the whole population of samples with which a comparison might be made if we knew the value of the troublesome parameter, but with a sub-population selected with reference to the sample in such a way that the distribution, in this sub-population, of the statistic used does not involve any unknown parameter. An example is the testing of significance of a regression coefficient. Thus if we suppose that a sample of values of  $x$  and  $y$  is drawn from a bivariate normal population, and calculate the regression coefficient  $b$  of  $y$  on  $x$  in the sample, the distribution of  $b$  involves not only the population value  $\beta$ , but also the ratio  $\alpha$  of the variances in the population. Since this second parameter is unknown, and can only be estimated from the sample, it is not possible to use the distribution of  $b$  in the whole population directly to test the significance of  $b - \beta$ . What we do is to find the place of this difference, not in the whole population of values in which both  $x$  and  $y$  are drawn at random, but in a sub-population for which the values of  $x$  are the same as in our sample. We may alternatively say that we limit the sub-population only to that for which the sum of the squares of the deviations of the values of  $x$  from their mean is the same as in our sample; the results are the same. The distribution in this sub-population of the ratio of  $b - \beta$  to its estimated standard error is of the Student form, with no unknown parameters, and on this basis it is possible to make exact and satisfactory tests and to set up fiducial limits for  $b$ . Another example is that of contingency tables. The practice now accepted (after a controversy) for testing independence of two modes of classification, such as classification of persons according as they have or have not been vaccinated, and again according as they live through an epidemic or die, is to compare the observed contingency table, not with all possible contingency tables of the same numbers of rows and columns, but only with the possible contingency tables having exactly the same marginal totals as the observed table.

**4. An Exact Solution.** We shall solve the problem of the significance of the difference of  $r_1$  and  $r_2$  with the understanding that the meaning of significance is to be interpreted by reference to the sub-population of possible samples for which the predictors  $x_1$  and  $x_2$  have the same set of values as those observed in the particular sample available. This procedure, besides yielding an exact distribution without unknown parameters, has the advantage of relaxing the stringency of the requirement of a trivariate normal distribution. We now make only the assumptions customary in the method of least squares, that the predictand  $y$  has the univariate normal distribution for each set of values of  $x_1$  and  $x_2$ , independently for the different sets, with a common variance  $\sigma^2$ , and with the expectation of  $y$  for a fixed pair of values of the predictors a linear function of these predictors. No assumption is involved regarding the distribution of the predictors, since we regard them as fixed in all the samples with which we compare our particular sample. The advantages of exactness and of freedom

from the somewhat special trivariate normal assumption are attained at the expense of sacrificing the precise applicability of the results to other sets of values of the predictors.

Since the correlational properties are unchanged by additive and multiplicative constants, we may suppose that

$$(1) \quad Sx_1 = 0 = Sx_2, \quad Sx_1^2 = 1 = Sx_2^2,$$

where  $S$  stands for summation over a sample of  $N$  individuals. The notation may be made more explicit by the adjunction of an additional subscript  $\alpha$ , varying from 1 to  $N$ , to denote the individual member of the sample, so that instead of  $Sx_1$ , for example, we might write  $Sx_{1\alpha}$ . The omission of this additional subscript is convenient and will usually leave no ambiguity when we deal with sums, but it will be convenient to retain it in connection with individual values. The correlation  $r_0$  of  $x_1$  with  $x_2$  in all those samples we shall consider is, by (1)

$$r_0 = Sx_1x_2.$$

Now consider the new quantities

$$(2) \quad x'_\alpha = \frac{x_{1\alpha} - x_{2\alpha}}{\sqrt{2(1 - r_0)}}, \quad x''_\alpha = \frac{x_{1\alpha} + x_{2\alpha}}{\sqrt{2(1 + r_0)}}.$$

Evidently, from (1) and (2),

$$(3) \quad Sx' = 0 = Sx'', \quad Sx'^2 = 1 = Sx''^2, \quad Sx'x'' = 0.$$

Since the mean value  $E(y_\alpha)$  is a linear function of  $x_{1\alpha}$  and  $x_{2\alpha}$ ,  $y_\alpha$  may, upon subtracting a constant from all these expectations, be written

$$(4) \quad y_\alpha = \beta_1 x_{1\alpha} + \beta_2 x_{2\alpha} + \Delta_\alpha,$$

where  $\Delta_1, \dots, \Delta_N$  are normally and independently distributed with variances all equal to  $\sigma^2$  and expectations zero. The assumption that  $x_1$  and  $x_2$  are equally correlated with  $y$  in the population leads to the conclusion that  $\beta_1 = \beta_2$ ; and putting  $\beta = \beta_1 \sqrt{2(1 + r_0)}$ , we then have from (4) and (2):

$$(5) \quad y_\alpha = \beta x''_\alpha + \Delta_\alpha.$$

Consequently, by (3)

$$Sx'y = Sx'_\alpha y_\alpha = \beta Sx'x'' + Sx'\Delta = Sx'\Delta;$$

and this function has a normal distribution with zero mean and variance  $\sigma^2$ .

If in the sample we work out a regression equation

$$Y = a + b'x' + b''x'',$$

the normal equations for determining  $b'$  and  $b''$  must by (3) take the simple forms

$$a = \bar{y}, \quad b' = Sx'y, \quad b'' = Sx''y.$$

From the general theory of least squares it is known that the sum of squares of residuals is

$$Sv^2 = S(y - \bar{y})^2 = Sy^2 - \bar{y}Sy - (Sx'y)^2 - S(x''y)^2,$$

and that  $Sv^2/\sigma^2$  has the  $\chi^2$  distribution with  $n = N - 3$  degrees of freedom, independently both of  $Sx'y$  and of  $Sx''y$ . From these facts it follows that

$$(6) \quad t = Sx'y \sqrt{\frac{n}{Sv^2}}$$

has the Student distribution with  $n$  degrees of freedom. Since in accordance with the foregoing definitions and (1) we have

$$Sx'y = (r_1 - r_2) \sqrt{\frac{S(y - \bar{y})^2}{2(1 - r_0)}},$$

and since also it is known that

$$Sv^2 = S(y - \bar{y})^2 \frac{D}{1 - r_0^2},$$

where

$$D = \begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_0 \\ r_2 & r_0 & 1 \end{vmatrix}$$

(6) may be written

$$(7) \quad t = (r_1 - r_2) \sqrt{\frac{n(1 + r_0)}{2D}}.$$

The probability of a greater value of  $|t|$  is given by tables of the Student distribution with  $n = N - 3$ . If this probability is sufficiently small (which conventionally means less than .05, or sometimes .01) we have a corresponding degree of confidence that the variate chosen because of a higher correlation in the sample has actually a higher correlation than the other in the population.

**5. The Selection of One Variate from Among Three or More.** Suppose that we are to choose one of the variates  $x_1, \dots, x_p$  in order to predict  $y$ . ( $p < N - 1$ ) We choose the one having highest correlation, and wonder how much confidence to place in this choice. We shall now determine the distribution of a function suitable for testing the hypothesis that there is no real difference between any pair of the correlations of  $x_1, \dots, x_p$  with  $y$ . Again we shall assume the values of these predictors fixed, and look for the place of our particular sample among all samples having these values, with only  $y$  free to vary normally by chance.

Let  $a_{ij} = S(x_i - \bar{x}_i)(x_j - \bar{x}_j)$ , and let  $c_{ij}$  be the cofactor of  $a_{ij}$  in the determinant  $a$  of these quantities, divided by  $a$ . Then

$$(8) \quad \sum a_{ij} c_{ik} = \delta_{jk} = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Here  $\Sigma$  stands for summation from 1 to  $p$ . Let

$$(9) \quad \frac{\sum c_{ij}}{\sum \sum c_{ij}},$$

$$(10) \quad l_i = S(x_i - \bar{x})y,$$

$$(11) \quad l = \Sigma w_i l_i.$$

From (9) it follows that

$$(12) \quad \Sigma w_i = 1.$$

From the hypothesis that  $y$  is in the population equally correlated with all the  $x_i$  it follows that  $l_1, \dots, l_p$  have equal expectations, which we may denote by  $\lambda$ ; and from (11) and (12) it follows that also  $E(l) = \lambda$ . Obviously

$$(13) \quad E(l_i - \lambda)(l_j - \lambda) = \sigma^2 a_{ij},$$

where  $\sigma^2$  is the variance of those values of  $y$  corresponding to a fixed set of values of the  $x$ 's. From (11), (13) and (9) we obtain

$$(14) \quad E(l - \lambda)^2 = \frac{\sigma^2}{\Sigma \Sigma c_{ij}}.$$

Since the  $l_i$  are linear functions of the  $y$ 's, they have the multivariate normal distribution. From the theory of this distribution and the values (13) of the covariances it follows that the distribution has the form

$$(2\pi)^{-1/2} a^{-1/2} \sigma^{-p} e^{-T/2\sigma^2} dl_1 \dots dl_p,$$

where  $a$  is the determinant of the  $a_{ij}$ 's, and

$$T = \Sigma \Sigma c_{ij} (l_i - \lambda)(l_j - \lambda).$$

We may introduce linear functions  $l'_1, \dots, l'_p$  of  $l_1 - \lambda, \dots, l_p - \lambda$  such that  $T = l'^2_1 + \dots + l'^2_p$ , and such that  $l'^2_p = (l - \lambda)^2 \Sigma \Sigma c_{ij}$ . Now  $\frac{l'^2_1 + \dots + l'^2_{p-1}}{\sigma^2}$  has the  $\chi^2$  distribution with  $p - 1$  degrees of freedom. The numerator of this expression equals

$$\begin{aligned} T - l'^2_p &= \Sigma \Sigma c_{ij} (l_i - \lambda)(l_j - \lambda) - (l - \lambda)^2 \Sigma \Sigma c_{ij} \\ &= \Sigma \Sigma c_{ij} l_i l_j - l^2 \Sigma \Sigma c_{ij} \\ &= \Sigma \Sigma c_{ij} (l_i - l)(l_j - l). \end{aligned}$$

The penultimate form shows that this function is independent of  $\lambda$ ; the last, as a positive definite form in the deviations of the  $l$ 's from their weighted mean, shows that sufficiently large values of the expression will reveal with definiteness the inequality of the predicting powers of the  $p$  variates when this exists.

It is well known that the regression coefficients of  $y$  upon the set of variates  $x_1, \dots, x_p$  are completely independent of the sum of squares  $Sv^2$  of residuals from the regression equation. Since the  $l$ 's are linear functions of these regression coefficients, (namely the linear functions appearing in the normal equations), they also are independent of  $Sv^2$ . Hence, if we put

$$s_1^2 = \frac{\sum \sum c_{ij} l_i l_j - l^2 \sum \sum c_{ij}}{p - 1}$$

$$s_2^2 = \frac{Sv^2}{N - p - 1},$$

the ratio  $F = s_1^2/s_2^2$  will, in case of equality of the correlations of the various  $x$ 's with  $y$ , have the variance ratio distribution with  $n_1 = p - 1$  and  $n_2 = N - p - 1$  degrees of freedom. When  $p = 2$  this test reduces exactly to (7), as it should, and  $F = t^2$ .

In the numerical application of this method, the regression coefficients  $b_i$  of  $y$  on  $x_1, \dots, x_p$  should first be worked out by the inverse matrix method. The right-hand members of the normal equations are  $l_1, \dots, l_p$ , the coefficients in these equations are the  $a_{ij}$ , and the calculation of  $s_1^2$  is simplified with the help of the identity

$$\sum \sum c_{ij} l_i l_j = \sum b_i l_i.$$

**6. Selection of Additional Variates When Some Have Been Chosen.** Suppose now that  $q$  predictors have been included definitely in the regression equation, and that one more is to be selected for inclusion among  $p$  additional predictors that are available. The criterion now is that that one should be chosen tentatively which has the highest partial correlation with the predictand, eliminating those already definitely chosen; but the confidence to be placed in the choice is to be judged by an adaptation of the criterion of the preceding section. It is only necessary to consider the  $a_{ij}$ ,  $l_i$ ,  $c_{ij}$  and  $b_i$  ( $i, j = 1, \dots, p$ ) as calculated from the new predictors and the deviations of  $y$  from the regression equation on the predictors already adopted. Formulae may easily be derived for the values of these quantities in terms of those already found and the sums of products, so as to simplify the calculations.  $Sv^2$  will now stand for the sum of squares of residuals from the regression equation involving all the  $p + q$  predictors. It is to be divided by  $N - p - q - 1$  to obtain  $s_2^2$ . The numbers of degrees of freedom with respect to which  $F$  is to be judged are now  $n_1 = p - 1$  and  $n_2 = N - p - q - 1$ . When  $p = 2$  this test, like that of the preceding section, reduces to the use of the  $t$ -distribution of (7), with  $n = N - q - 3$ , and the correlations standing for partial correlations eliminating the predictors already definitely chosen.

A special instance in which this procedure is applicable is in economic time series, in which time, in the form of orthogonal polynomials, must ordinarily be "partialled out" in order that tests of significance may be sound.

**7. Further Problems.** It is natural to ask whether the foregoing work can be extended to examine the soundness of the selection, on the basis of a greater multiple correlation, of a particular set of two or more variates, chosen from among several such sets. The simplest such problem that goes beyond what has been done above deals with two sets, each of two predictors, having in a sample multiple correlations  $R$  and  $R'$  with the predictand. The question is whether the difference  $R - R'$  is significant.

Suppose that, in the interests of simplicity and the hope of attaining a solution satisfactorily free from unknown parameters, we assume as before that the predictors have a fixed set of values, the same in all samples. Since multiple correlations are invariant under linear transformations of predictors, we may without loss of generality assume that the predictors in each set are mutually uncorrelated and have sums of squares equal to unity. Indeed, we may go somewhat further in standardizing the sets of values to which consideration can be confined without loss of generality, with the help of some ideas introduced in the paper [1]. In the terminology of that paper, the variates in each set may be considered *canonical* with respect to the relationship between the sets. This means that linear functions  $x_1$  and  $x_2$  of the two variates in one set, and linear functions  $x'_1$  and  $x'_2$  of those in the other set, can be chosen so as to satisfy not only the conditions

$$\begin{aligned} Sx_1 &= Sx_2 = Sx'_1 = Sx'_2 = 0 \\ (15) \quad Sx_1^2 &= Sx_2^2 = Sx_1'^2 = Sx_2'^2 = 1 \\ Sx_1x_2 &= 0 = Sx'_1x'_2, \end{aligned}$$

but also the further conditions

$$(16) \quad Sx_1x'_2 = 0 = Sx_2x'_1.$$

This means that, for all the purposes in view, the two sets of predictors can be characterized as to their mutual relationships by the values of the remaining two sums of products, namely

$$c_1 = Sx_1x'_1, \quad c_2 = Sx_2x'_2.$$

In view of the conditions assumed earlier,  $c_1$  and  $c_2$  are what have been called the *canonical correlations* between the two sets.

To the sets thus standardized, the predictand  $y$  is related in a manner expressed by the population regression coefficients  $\beta_1$  and  $\beta_2$  of  $y$  on the first set, and  $\beta'_1$  and  $\beta'_2$  on the second. If we take  $y$  as having unit variance in the population, the squared multiple correlation coefficients in the two cases will be

$$\rho^2 = \beta_1^2 + \beta_2^2, \quad \rho'^2 = \beta_1'^2 + \beta_2'^2.$$

The hypothesis to be tested is that  $\rho = \rho'$ . If  $b_1, b_2, b'_1, b'_2$  denote the sample estimates of the regression coefficients, the statistic appropriate for the test would appear necessarily to be proportional to

$$w = \frac{1}{2}(b_1^2 + b_2^2 - b_1'^2 - b_2'^2).$$



The sample regression coefficients are normally distributed, with population correlations equal to the sample correlations among the corresponding predictors. The variance of each is  $\sigma^2$ . Thus their joint distribution may be written down at once, in a rather simple form in view of (15) and (16). From this it is possible to determine directly the characteristic function  $M(t) = Ee^{tw}$  of  $w$ . If we write  $K(t) = \log M(t)$  we obtain:

$$2K(t) = \Sigma \{ (\beta_j^2 - 2c_j\beta_j\beta'_j + \beta_j'^2)t^2 + (\beta_j^2 - \beta_j'^2)t \} \{ 1 - (1 - c_j^2)t^2 \}^{-1} \\ - \Sigma \log \{ 1 - (1 - c_j^2)t^2 \}.$$

Here the summations are with respect to  $j$  over the values 1 and 2. If each set of predictors had had  $s$  members, the same result would hold for  $K(t)$  except that the summations with respect to  $j$  would then extend from 1 to  $s$ .

This is a very disappointing result because it contains so many parameters. The distribution of  $w$  must contain the same parameters as its characteristic function. All the four parameters  $\beta_j, \beta'_j$  appear in the expression above, though their effective number is reduced to three by the condition that the two sums of squares shall be equal which constitutes the hypothesis under test. The distribution of  $w$  thus contains at least three unknown parameters besides  $\sigma$ .

The estimate of variance  $s^2$  obtained from the residuals from the grand regression equation of  $y$  on  $x_1, x_2, x'_1$ , and  $x'_2$  is independent of  $w$ . Its distribution is of the usual form and involves a parameter, the population variance, which is a function of  $\beta_1, \beta_2, \beta'_1$ , and  $\beta'_2$ . We could therefore pass by a single integration from the distribution of  $w$  to that of the statistic  $w/s^2$ , which vanishes with  $w$ , and which on this account, and on grounds of physical dimensionality, might be considered appropriate to test the hypothesis that  $\rho = \rho'$ . The question may be raised whether the distribution of this ratio might not be free from parameters. The answer unfortunately is in the negative, as appears from an examination of the characteristic function of the ratio. Even in the simplified case in which all the  $c_j$  are equal, a troublesome parameter persists in the distribution.

Thus we meet again the problem of nuisance parameters, and this time no escape is visible. Perhaps some such artifice as those enumerated in paragraph 3 (for example, some further limitation of the sub-population within which we should seek the place of our particular sample) is capable of yielding an exact, or "studentized" distribution, but this has not yet been found. The problem is of considerable interest, not only because of its practical importance, but because of its suggestiveness in connection with general theory.

Numerous other problems having both practical importance and general theoretical interest are associated with the selection of predictors. For example, we have not dealt at all with the problem of the *number* of predictors that should be used when maximum accuracy in prediction, or in evaluation of the regression coefficients, is the sole criterion. A particular case is the determination of the degree of the regression polynomial which should be fitted to obtain

maximum accuracy, for example of the number of orthogonal polynomials in fitting a trend. Such customary criteria as minimizing the *estimated* variance of deviations, in which the sum of squares which is the numerator and the number of degrees of freedom which is the denominator both diminish to zero as the number of variates is increased, do not rest upon any satisfactory general theory.

Another related set of problems is concerned with variates more numerous than the observations on each. It is clear that there is real information inherent in data of this kind, but existing theory and methods, including those of the present paper, are not adequate to utilize it in a thoroughly efficient manner. A recent paper of P. L. Hsu [9] is unique in not excluding the case in which the variates outnumber the observations.

**8. Summary.** A criterion has been obtained for judging the definiteness of the selection of a particular variate, from among several available for prediction, on the basis of its having the maximum sample correlation with the predictand. A variation of this criterion is applied in paragraph 6 to the problem of extending the list of variates to be used in a regression formula.

Some of the problems of "nuisance parameters" which affect general theory are illustrated in this problem. Some outstanding unsolved problems related to these questions are discussed in paragraph 7.

#### REFERENCES

- [1] Harold Hotelling, "Relations Between Two Sets of Variates," *Biometrika*, Vol. 28 (1936), pp. 321-377.
- [2] "Student," "The Probable Error of a Mean," *Biometrika*, Vol. 6 (1908), pp. 1-25.
- [3] R. A. Fisher, "Applications of 'Student's' Distribution," *Metron*, Vol. 5 (1925), pp. 90-104.
- [4] R. A. Fisher, "The Fiducial Argument in Statistical Inference," *Annals of Eugenics*, Vol. 6 (1935), pp. 391-398. See also Fisher's answer to Bartlett in the *Annals of Math. Stat.*, Vol. 10 (1939), pp. 383-388 and the references there given.
- [5] Daisy M. Starkey, "A Test of the Significance of the Difference Between Means of Samples from Two Normal Populations Without Assuming Equal Variances," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 201-213.
- [6] M. S. Bartlett, "The Information Available in Small Samples," *Proc. Camb. Phil. Soc.*, Vol. 32 (1936), pp. 560-566.
- [7] B. L. Welch, "On Confidence Limits and Sufficiency, with Particular Reference to Parameters of Location," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 58-69.
- [8] *Statistical Research Memoirs*, Vol. 2 (1938), pp. 58-59.
- [9] P. L. Hsu, "On the Distribution of Roots of Certain Determinantal Equations," *Annals of Eugenics*, Vol. 9 (1939), pp. 250-258.

COLUMBIA UNIVERSITY,  
NEW YORK, N. Y.

## THE FITTING OF STRAIGHT LINES IF BOTH VARIABLES ARE SUBJECT TO ERROR

BY ABRAHAM WALD

**1. Introduction.** The problem of fitting straight lines if both variables  $x$  and  $y$  are subject to error, has been treated by many authors. If we have  $N > 2$  observed points  $(x_i, y_i)$  ( $i = 1, \dots, N$ ), the usually employed method of least squares for determining the coefficients  $a, b$ , of the straight line  $y = ax + b$  is that of choosing values of  $a$  and  $b$  which minimize the sum of the squares of the residuals of the  $y$ 's, i.e.  $\sum(ax_i + b - y_i)^2$  is a minimum. It is well known that treating  $y$  as an independent variable and minimizing the sum of the squares of the residuals of the  $x$ 's, we get a different straight line as best fit. It has been pointed out<sup>1</sup> that if both variables are subject to error there is no reason to prefer one of the regression lines described above to the other. For obtaining the "best fit," which is not necessarily equal to one of the two lines mentioned, new criteria have to be found. This problem was treated by R. J. Adcock as early as 1877.<sup>2</sup>

He defines the line of best fit as the one for which the sum of the squares of the normal deviates of the  $N$  observed points from the line becomes a minimum. (Another early attempt to solve this problem by minimizing the sum of squares of the normal deviates was made by Karl Pearson.<sup>3</sup>)

Many objections can be raised against this method. First, there is no justification for minimizing the sum of the squares of the *normal* deviates, and not the deviations in some other direction. Second, the straight line obtained by that method is not invariant under transformation of the coordinate system. It is clear that a satisfactory method should give results which do not depend on the choice of a particular coordinate system. This point has been emphasized by C. F. Roos. He gives<sup>4</sup> a good summary of the different methods and then proposes a general formula for fitting lines (and planes in case of more than two variables) which do not depend on the choice of the coordinate system.

<sup>1</sup> See for instance Henry Schultz' "The Statistical Law of Demand," *Jour. of Political Economy*, Vol. 33, Dec. (1925).

<sup>2</sup> *Analyst*, Vol. IV, p. 183 and Vol. V, p. 53.

<sup>3</sup> "On Lines and Planes of Closest Fit to Systems of Points in Space" *Phil. Mag.* 6th Ser. Vol. II (1901).

<sup>4</sup> "A General Invariant Criterion of Fit for Lines and Planes where all Variates are Subject to Error," *Metron*, February 1937. See also Oppenheim and Roos *Bulletin of the American Mathematical Society*, Vol. 34 (1928), pp. 140-141.

Roos' formula includes many previous solutions<sup>5</sup> as special cases. H. E. Jones<sup>6</sup> gives an interesting geometric interpretation of Roos' general formula.

It is a common feature of Roos' general formula and of all other methods proposed in recent years that the fitted straight line cannot be determined without *a priori* assumptions (independent of the observations) regarding the weights of the errors in the variables  $x$  and  $y$ . That is to say, either the standard deviations of the errors in  $x$  and in  $y$  are involved (or at least their ratio is included) in the formula of the fitted straight line and there is no method given by which those standard deviations can be estimated by means of the observed values of  $x$  and  $y$ .

R. Frisch<sup>7</sup> has developed a new general theory of linear regression analysis, when all variables are subject to error. His very interesting theory employs quite new methods and is not based on probability concepts. Also on the basis of Frisch's discussion it seems that there is no way of determining the "true" regression without *a priori* assumptions about the disturbing intensities.

T. Koopmans<sup>8</sup> combined Frisch's regression theory with the classical one in a new general theory based on probability concepts. Also, according to his theory, the regression line can be determined only if the ratio of the standard deviations of the errors is known.

In a recent paper R. G. D. Allen<sup>9</sup> gives a new interesting method for determining the fitted straight line in case of two variables  $x$  and  $y$ . Denoting by  $\sigma_x$  the standard deviation of the errors in  $x$ , by  $\sigma_y$  the standard deviation of the errors in  $y$  and by  $\rho$  the correlation coefficient between the errors in the two variables, Allen emphasizes (p. 194)<sup>9</sup> that the fitted line can be determined only if the values of two of the three quantities  $\sigma_x$ ,  $\sigma_y$ ,  $\rho$  are given *a priori*.

Finally I should like to mention a paper by C. Eisenhart,<sup>10</sup> which contains many interesting remarks related to the subject treated here.

In the present paper I shall deal with the case of two variables  $x$  and  $y$  in which the errors are uncorrelated. It will be shown that under certain conditions:

- (1) The fitted straight line can be determined without making *a priori* assumptions (independent of the observed values  $x$  and  $y$ ) regarding the standard deviations of the errors.
- (2) The standard deviation of the errors can be well estimated by means of

<sup>5</sup> For instance also Corrado Gini's method described in his paper, "Sull' Interpolazione di una Retta Quando i Valori della Variable Indipendente sono Affetti da Errori Accidentalit ," *Metron*, Vol. I, No. 3 (1921), pp. 63-82.

<sup>6</sup> "Some Geometrical Considerations in the General Theory of Fitting Lines and Planes," *Metron*, February 1937.

<sup>7</sup> *Statistical Confluence Analysis by Means of Complete Regression Systems*, Oslo, 1934.

<sup>8</sup> *Linear Regression Analysis of Economic Time Series*, Haarlem, 1937.

<sup>9</sup> "The Assumptions of Linear Regression," *Economica*, May 1939.

<sup>10</sup> "The interpretation of certain regression methods and their use in biological and industrial research," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 162-186.

the observed values of  $x$  and  $y$ . The precision of the estimate increases with the number of the observations and would give the exact values if the number of observations were infinite. (See in this connection also condition V in section 3.)

**2. Formulation of the Problem.** Let us begin with a precise formulation of the problem. We consider two sets of random variables<sup>11</sup>

$$x_1, \dots, x_N; \quad y_1, \dots, y_N.$$

Denote the expected value  $E(x_i)$  of  $x_i$  by  $X_i$  and the expected value  $E(y_i)$  of  $y_i$  by  $Y_i$  ( $i = 1, \dots, N$ ). We shall call  $X_i$  the true value of  $x_i$ ,  $Y_i$  the true value of  $y_i$ ,  $x_i - X_i = \epsilon_i$  the error in the  $i$ -th term of the  $x$ -set, and  $y_i - Y_i = \eta_i$  the error in the  $i$ -th term of the  $y$ -set.

The following assumptions will be made:

I. The random variables  $\epsilon_1, \dots, \epsilon_N$  each have the same distribution and they are uncorrelated, i.e.  $E(\epsilon_i \epsilon_j) = 0$  for  $i \neq j$ . The variance of  $\epsilon_i$  is finite.

II. The random variables  $\eta_1, \dots, \eta_N$  each have the same distribution and are uncorrelated, i.e.  $E(\eta_i \eta_j) = 0$  for  $i \neq j$ . The variance of  $\eta_i$  is finite.

III. The random variables  $\epsilon_i$  and  $\eta_j$  ( $i = 1, \dots, N; j = 1, \dots, N$ ) are uncorrelated, i.e.  $E(\epsilon_i \eta_j) = 0$ .

IV. A single linear relation holds between the true values  $X$  and  $Y$ , that is to say  $Y_i = \alpha X_i + \beta$  ( $i = 1, \dots, N$ ).

Denote by  $\epsilon$  a random variable having the same probability distribution as possessed by each of the random variables  $\epsilon_1, \dots, \epsilon_N$ , and by  $\eta$  a random variable having the same distribution as  $\eta_1, \dots, \eta_N$ .

The problem to be solved can be formulated as follows:

We know only two sets of observations:  $x'_1, \dots, x'_N; y'_1, \dots, y'_N$ , where  $x'_i$  denotes the observed value of  $x_i$  and  $y'_i$  denotes the observed value of  $y_i$ . We know neither the true values  $X_1, \dots, X_N; Y_1, \dots, Y_N$ , nor the coefficients  $\alpha$  and  $\beta$  of the linear relation between them. We have to estimate by means of the observations  $x'_1, \dots, x'_N; y'_1, \dots, y'_N$ , (1) the values of  $\alpha$  and  $\beta$ , (2) the standard deviation  $\sigma_\epsilon$  of  $\epsilon$ , and (3) the standard deviation  $\sigma_\eta$  of  $\eta$ .

Problems of this kind occur often in Economics, where we are dealing with time series. For example, denote by  $x_i$  the price of a certain good  $G$  in the period  $t_i$ , and by  $y_i$  the quantity of  $G$  demanded in  $t_i$ . In each time period  $t_i$  there exists a normal price  $X_i$  and a normal demand  $Y_i$  which would obtain if the influence of some accidental disturbances could be eliminated. If we have reason to assume that there exists between the normal price and the normal demand a linear relationship we have to deal with a problem of the kind described above.

In the following discussions we shall use the notations  $x_i$  and  $y_i$  also for their

<sup>11</sup> A random or stochastic variable is a real variable associated with a probability distribution.

observed values  $x'_i$  and  $y'_i$  since it will be clear in which sense they are meant and no confusion can arise.

**3. Consistent Estimates of the Parameters  $\alpha$ ,  $\beta$ ,  $\sigma_\epsilon$ ,  $\sigma_\eta$ .** For the sake of simplicity we assume that  $N$  is even. We consider the expression

$$(1) \quad \begin{aligned} a_1 &= \frac{(x_1 + \dots + x_m) - (x_{m+1} + \dots + x_N)}{N} \\ a_2 &= \frac{(y_1 + \dots + y_m) - (y_{m+1} + \dots + y_N)}{N}, \end{aligned}$$

where  $m = N/2$ . As an estimate of  $\alpha$  we shall use the expression

$$(2) \quad a = \frac{a_2}{a_1} = \frac{(y_1 + \dots + y_m) - (y_{m+1} + \dots + y_N)}{(x_1 + \dots + x_m) - (x_{m+1} + \dots + x_N)}.$$

We make the assumption

V. *The limit inferior of*

$$\left| \frac{(X_1 + \dots + X_m) - (X_{m+1} + \dots + X_N)}{N} \right| \quad (N = 2, 3, \dots \text{ad. inf.})$$

is positive.

We shall prove that  $a$  is a consistent estimate of  $\alpha$ , i.e.  $a$  converges stochastically to  $\alpha$  with  $N \rightarrow \infty$ , if the assumptions I-V hold. Denote the expected value of  $a_1$  by  $\bar{a}_1$  and the expected value of  $a_2$  by  $\bar{a}_2$ . It is obvious that

$$(3) \quad \begin{aligned} \bar{a}_1 &= \frac{(X_1 + \dots + X_m) - (X_{m+1} + \dots + X_N)}{N} \\ \bar{a}_2 &= \frac{(Y_1 + \dots + Y_m) - (Y_{m+1} + \dots + Y_N)}{N}. \end{aligned}$$

On account of the condition IV we have

$$(4) \quad \bar{a}_2 = \alpha \bar{a}_1, \quad \text{or} \quad \frac{\bar{a}_2}{\bar{a}_1} = \alpha.$$

The variance of  $a_1 - \bar{a}_1$  is equal to  $\sigma_\epsilon^2/N$  and the variance of  $a_2 - \bar{a}_2$  is equal to  $\sigma_\eta^2/N$ . Hence  $a_1$  and  $a_2$  converge stochastically towards  $\bar{a}_1$  and  $\bar{a}_2$  respectively. From that and assumption V it follows that also  $\frac{a_2}{a_1}$  converges stochastically towards  $\frac{\bar{a}_2}{\bar{a}_1} = \alpha$ . The intercept  $\beta$  of the regression line will be estimated by

$$(5) \quad b = \bar{y} - \alpha \bar{x}, \quad \text{where } \bar{x} = \frac{x_1 + \dots + x_N}{N} \quad \text{and} \quad \bar{y} = \frac{y_1 + \dots + y_N}{N}.$$

Denote by  $\bar{X}$  the arithmetic mean of  $X_1, \dots, X_N$  and by  $\bar{Y}$  the arithmetic mean of  $Y_1, \dots, Y_N$ . Since  $\bar{y}$  converges stochastically towards  $\bar{Y}$ ,  $\bar{x}$  towards

$\bar{X}$ , and  $a$  towards  $\alpha$ ,  $b$  converges stochastically towards  $\bar{Y} - \alpha\bar{X}$ . From condition IV it follows that  $\bar{Y} - \alpha\bar{X} = \beta$ . Hence  $b$  converges stochastically towards  $\beta$ .

Let us introduce the following notations:

$s_x = \sqrt{\sum \frac{(x_i - \bar{x})^2}{N}}$  = sample standard deviation of the  $x$ -observations,

$s_y = \sqrt{\sum \frac{(y_i - \bar{y})^2}{N}}$  = sample standard deviation of the  $y$ -observations,

$s_{xy} = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{N}$  = sample covariance between the  $x$ -set and  $y$ -set.

$s_x$ ,  $s_y$  and  $s_{xy}$  denote the same expressions of the true values  $X_1, \dots, X_N$ ;  $Y_1, \dots, Y_N$ .

It is obvious that

$$(6) \quad E(s_x^2) = s_x^2 + \sigma_\epsilon^2 \frac{N-1}{N},$$

$$(7) \quad E(s_y^2) = s_y^2 + \sigma_\eta^2 \frac{N-1}{N},$$

$$(8) \quad E(s_{xy}) = s_{xy},$$

where  $E(s_x^2)$ ,  $E(s_y^2)$ , and  $E(s_{xy})$  denote the expected values of  $s_x^2$ ,  $s_y^2$ , and  $s_{xy}$ .<sup>12</sup>

Since  $Y_i = \alpha X_i + \beta$ , we have

$$(9) \quad s_y = \alpha s_x,$$

$$(10) \quad s_{xy} = \alpha s_x^2.$$

From (8), (9) and (10) we get

$$(11) \quad s_x^2 = \frac{E(s_{xy})}{\alpha},$$

$$(12) \quad s_y^2 = \alpha E(s_{xy}).$$

If we substitute in (6) and (7) for  $s_x^2$  and  $s_y^2$  their values in (11) and (12), we get

$$(13) \quad \sigma_\epsilon^2 = \left[ E(s_x^2) - \frac{E(s_{xy})}{\alpha} \right] N / (N-1),$$

$$(14) \quad \sigma_\eta^2 = [E(s_y^2) - \alpha E(s_{xy})] N / (N-1).$$

<sup>12</sup> I observe that the equations (6), (7) and (8) are essentially the same as those investigated by R. Frisch, *Statistical Confluence Analysis* pp. 51-52. See also Allen's equations (4) l.c. p. 194.

Since  $s_x^2$ ,  $s_y^2$ ,  $s_{xy}$  converge stochastically towards their expected values and  $a$  converges stochastically towards  $\alpha$ , the expressions

$$(15) \quad \left[ s_x^2 - \frac{s_{xy}^2}{a} \right] N / (N - 1)$$

and

$$(16) \quad [s_y^2 - a s_{xy}] N / (N - 1)$$

are consistent estimates of  $\sigma_x^2$  and  $\sigma_y^2$  respectively.

**4. Confidence Interval for  $\alpha$ .** In this section, as well as in sections 5 and 6, only the assumptions I-IV are assumed to hold. In other words, all statements made in these sections are valid independently of Assumption V, except where the contrary is explicitly stated.

Let us introduce the following notation:

$$\begin{aligned} \bar{x}_1 &= \frac{x_1 + \dots + x_m}{m}; & \bar{y}_1 &= \frac{y_1 + \dots + y_m}{m} \\ \bar{x}_2 &= \frac{x_{m+1} + \dots + x_N}{m}; & \bar{y}_2 &= \frac{y_{m+1} + \dots + y_N}{m} \\ (s'_x)^2 &= \frac{\sum_{i=1}^m (x_i - \bar{x}_1)^2 + \sum_{j=m+1}^N (x_j - \bar{x}_2)^2}{N} \\ (s'_y)^2 &= \frac{\sum_{i=1}^m (y_i - \bar{y}_1)^2 + \sum_{j=m+1}^N (y_j - \bar{y}_2)^2}{N} \\ s'_{xy} &= \frac{\sum_{i=1}^m (x_i - \bar{x}_1)(y_i - \bar{y}_1) + \sum_{j=m+1}^N (x_j - \bar{x}_2)(y_j - \bar{y}_2)}{N}. \end{aligned}$$

$\bar{X}_1$ ,  $\bar{X}_2$ ,  $\bar{Y}_1$ ,  $\bar{Y}_2$ ,  $(s'_x)^2$ ,  $(s'_y)^2$  and  $s'_{xy}$  denote the same functions of the true values  $X_1, \dots, X_N$ ,  $Y_1, \dots, Y_N$ . The expressions  $s'_x$ ,  $s'_y$ , and  $s'_{xy}$  are slightly different from the corresponding expressions  $s_x$ ,  $s_y$ , and  $s_{xy}$ . The reason for introducing these new expressions is that the distributions of  $s_x$ ,  $s_y$ , and  $s_{xy}$  are not independent of the slope  $a = \frac{a_2}{a_1}$  of the sample regression line, but  $s'_x$ ,  $s'_y$  and  $s'_{xy}$  are distributed independently from  $a$  (assuming that  $\epsilon$  and  $\eta$  are normally distributed). The latter statement follows easily from the fact that according to (1) and (2)  $a = \frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2}$  and  $s'_x$ ,  $s'_y$ ,  $s'_{xy}$  are distributed independently of  $\bar{x}_1$ ,  $\bar{x}_2$ ,  $\bar{y}_1$  and  $\bar{y}_2$ .



In the same way as we derived (13) and (14), we get

$$(13') \quad \sigma_\epsilon^2 = \left[ E(s_x')^2 - \frac{E(s_{xy}')^2}{\alpha} \right] N / (N - 2),$$

$$(14') \quad \sigma_\eta^2 = [E(s_y')^2 - \alpha E(s_{xy}')^2] N / (N - 2).$$

These formulae differ from the corresponding formulae (13) and (14) only in the denominator of the second factor, having there  $N - 2$  instead of  $N - 1$ . This is due to the fact that the estimates  $s_x$ ,  $s_y$ ,  $s_{xy}$  are based on  $N - 1$  degrees of freedom whereas  $s_x'$ ,  $s_y'$  and  $s_{xy}'$  are based only on  $N - 2$  degrees of freedom. From (13') and (14') we get the following estimates<sup>13</sup> for  $\sigma_\epsilon^2$  and  $\sigma_\eta^2$ :

$$(17) \quad \left[ (s_x')^2 - \frac{s_{xy}'^2}{\alpha} \right] N / (N - 2),$$

$$(18) \quad [(s_y')^2 - \alpha s_{xy}'^2] N / (N - 2).$$

Hence we get as an estimate of  $\sigma_\eta^2 + \alpha^2 \sigma_\epsilon^2$  the expression:

$$(19) \quad s^2 = [(s_y')^2 + \alpha^2 (s_x')^2 - 2\alpha s_{xy}'] N / (N - 2) \\ = \frac{N}{N - 2} \left\{ \frac{\sum_{i=1}^m [(y_i - \alpha x_i) - (\bar{y}_1 - \alpha \bar{x}_1)]^2}{m} + \frac{\sum_{j=m+1}^N [(y_j - \alpha x_j) - (\bar{y}_2 - \alpha \bar{x}_2)]^2}{N} \right\}$$

Now we shall show that

$$(20) \quad \frac{(N - 2)s^2}{\sigma_\eta^2 + \alpha^2 \sigma_\epsilon^2}$$

has the  $\chi^2$ -distribution with  $N - 2$  degrees of freedom, provided that  $\epsilon$  and  $\eta$  are normally distributed. In fact,

$$(y_i - \alpha x_i) - (\bar{y}_1 - \alpha \bar{x}_1) = \eta_i - \alpha \epsilon_i - (\bar{\eta}_1 - \alpha \bar{\epsilon}_1) \quad (i = 1, \dots, m)$$

and

$$(y_j - \alpha x_j) - (\bar{y}_2 - \alpha \bar{x}_2) = \eta_j - \alpha \epsilon_j - (\bar{\eta}_2 - \alpha \bar{\epsilon}_2) \quad (j = m + 1, \dots, N),$$

where

$$\bar{\epsilon}_1 = \frac{\epsilon_1 + \dots + \epsilon_m}{m}, \quad \bar{\epsilon}_2 = \frac{\epsilon_{m+1} + \dots + \epsilon_N}{m}, \\ \bar{\eta}_1 = \frac{\eta_1 + \dots + \eta_m}{m}, \quad \bar{\eta}_2 = \frac{\eta_{m+1} + \dots + \eta_N}{m}.$$

Since the variance of  $\eta_k - \alpha \epsilon_k$  is equal to  $\sigma_\eta^2 + \alpha^2 \sigma_\epsilon^2$  and since  $\eta_k - \alpha \epsilon_k$  is uncorrelated with  $\eta_l - \alpha \epsilon_l$  ( $k \neq l$ ) ( $k, l = 1, \dots, N$ ), the expression (20) has the  $\chi^2$ -distribution with  $N - 2$  degrees of freedom.

<sup>13</sup> An "estimate" is usually a function of the observations not involving any unknown parameters. We designate here as estimates also some functions involving the parameter  $\alpha$ .

Now we shall show that

$$(21) \quad \frac{\sqrt{N} a_1(a - \alpha)}{\sqrt{\sigma_\eta^2 + \alpha^2 \sigma_\epsilon^2}}$$

is normally distributed with zero mean and unit variance. In fact from the equations (1)-(4) it follows that

$$\begin{aligned} a_1(a - \alpha) &= \bar{a}_2 + \frac{\eta_1 - \eta_2}{2} - a_1 \left( \frac{\bar{a}_2}{\bar{a}_1} \right) \\ &= \bar{a}_2 + \frac{\eta_1 - \eta_2}{2} - \left( \bar{a}_1 + \frac{\bar{\epsilon}_1 - \bar{\epsilon}_2}{2} \right) \left( \frac{\bar{a}_2}{\bar{a}_1} \right) \\ &= \frac{\eta_1 - \eta_2}{2} - \alpha \frac{\bar{\epsilon}_1 - \bar{\epsilon}_2}{2}. \end{aligned}$$

Since the latter expression is normally distributed (provided that  $\epsilon$  and  $\eta$  are normally distributed) with zero mean and variance  $\frac{\sigma_\eta^2 + \alpha^2 \sigma_\epsilon^2}{N}$ , our statement about (21) is proved.

Obviously (20) and (21) are independently distributed, hence  $\sqrt{N-2}$  times the ratio of (21) to the square root of (20), namely,

$$(22) \quad t = \sqrt{N-2} \frac{\sqrt{N} a_1(a - \alpha)}{\sqrt{N-2} s} = \frac{a_1(a - \alpha) \sqrt{N-2}}{\sqrt{(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}}}$$

has the Student distribution with  $N-2$  degrees of freedom. Denote by  $t_0$  the critical value of  $t$  corresponding to a chosen probability level. The deviation of  $a$  from an assumed population value  $\alpha$  is significant if

$$\left| \frac{a_1(a - \alpha) \sqrt{N-2}}{\sqrt{(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}}} \right| \geq t_0.$$

The confidence interval for  $\alpha$  can be obtained by solving the equation in  $\alpha$ ,

$$(23) \quad a_1^2(a - \alpha)^2 = [(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}] \frac{t_0^2}{N-2}.$$

Now we shall show that if the relation

$$(24) \quad a_1^2 > \frac{(s'_x)^2 t_0^2}{N-2},$$

holds, the roots  $\alpha_1$  and  $\alpha_2$  are real and  $a$  is contained in the interior of the interval  $[\alpha_1, \alpha_2]$ . From (19) it follows that

$$(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy} > 0$$

for all values of  $\alpha$ . Hence, for  $\alpha = a$  the left hand side of (23) is smaller than the right hand side. On account of (24) there exists a value  $a' > a$  and a

value  $a'' < a$  such that the left hand side of (23) is greater than the right hand side for  $\alpha = a'$  and  $\alpha = a''$ . Hence one root must lie between  $a$  and  $a'$  and the other root between  $a''$  and  $a$ . This proves our statement. The relation (24) always holds for sufficiently large  $N$  if Assumption V is fulfilled. The confidence interval of  $\alpha$  is the interval  $[\alpha_1, \alpha_2]$ . For very small  $N$  (24) may not hold.

Finally I should like to remark that no essentially better estimate of the variance of  $\eta - \alpha\epsilon$  can be given than the expression  $s^2$  in (19). In fact, we have  $2N$  observations  $x_1, \dots, x_N; y_1, \dots, y_N$ . For the estimation of the variance of  $\eta - \alpha\epsilon$  we must eliminate the unknowns  $X_1, \dots, X_N$  and  $\beta$ . (The unknowns  $Y_1, \dots, Y_N$  are determined by the relations  $Y_i = \alpha X_i + \beta$  and  $\alpha$  is involved in the expression whose variance is to be determined.) Hence we have at most  $N - 1$  degrees of freedom and the estimate in (19) is based on  $N - 2$  degrees of freedom.

**5. Confidence Interval for  $\beta$  if  $\alpha$  is Given.** In this case the best estimate of  $\beta$  is given by the expression:

$$b_\alpha = \bar{y} - \alpha \bar{x} \text{ where } \bar{x} = \frac{x_1 + \dots + x_N}{N} \text{ and } \bar{y} = \frac{y_1 + \dots + y_N}{N}.$$

We have

$$b_\alpha - \beta = (\bar{y} - \bar{Y}) - \alpha(\bar{x} - \bar{X}) = \eta - \alpha\bar{\epsilon}$$

where

$$\bar{\epsilon} = \frac{\epsilon_1 + \dots + \epsilon_N}{N}, \text{ and } \eta = \frac{\eta_1 + \dots + \eta_N}{N}.$$

Hence,

$$(25) \quad \frac{\sqrt{N} (b_\alpha - \beta)}{\sqrt{\sigma_\eta^2 + \alpha^2 \sigma_\epsilon^2}}$$

is normally distributed with zero mean and unit variance. It is obvious that the expressions (20) and (25) are independently distributed. Hence  $\sqrt{N - 2}$  times the ratio of (25) to the square root of (20), i.e.

$$t = \sqrt{N - 2} \frac{\sqrt{N} (b_\alpha - \beta)}{\sqrt{N - 2} s} = \frac{\sqrt{N - 2} (b_\alpha - \beta)}{\sqrt{(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_x s'_y}}$$

has the Student distribution with  $N - 2$  degrees of freedom. Denoting by  $t_0$  the critical value of  $t$  according to the chosen probability level, the confidence interval for  $\beta$  is given by the interval:

$$\left[ b_\alpha + \frac{\sqrt{(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_x s'_y}}{\sqrt{N - 2}} t_0, \quad b_\alpha - \frac{\sqrt{(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_x s'_y}}{\sqrt{N - 2}} t_0 \right].$$

**6. Confidence Region for  $\alpha$  and  $\beta$  Jointly.** In most practical cases we want to know confidence limits for  $\alpha$  and  $\beta$  jointly. A pair of values  $\alpha, \beta$  can be represented in the plane by the point with the coordinates  $\alpha, \beta$ . A region  $R$  of this plane is called confidence region of the true point  $(\alpha, \beta)$  corresponding to the probability level  $P$  if the following two conditions are fulfilled.

(1) The region  $R$  is a function of the observations  $x_1, \dots, x_N; y_1, \dots, y_N$ , i.e. it is uniquely determined by the observations.

(2) Before performing the experiment the probability that we shall obtain observed values such that  $(\alpha, \beta)$  will be contained in  $R$ , is exactly equal to  $P$ .  $P$  is usually chosen to be equal to .95 or .99.

We have shown that the expressions (21) and (25), i.e.

$$\frac{\sqrt{N} a_1(a - \alpha)}{\sqrt{\sigma_y^2 + \alpha^2 \sigma_x^2}}, \quad \frac{\sqrt{N} (b_a - \beta)}{\sqrt{\sigma_y^2 + \alpha^2 \sigma_x^2}}$$

are normally distributed with zero mean and unit variance. Now we shall show that these two quantities are independently distributed. For this purpose we have only to show that  $\bar{x}, \bar{y}, a_1$  and  $a_2$  are independently distributed ( $a_1$  and  $a_2$  are defined in (1)), but since

$$a_1 - E(a_1) = (\bar{\epsilon}_1 - \bar{\epsilon}_2)/2$$

$$a_2 - E(a_2) = (\eta_1 - \eta_2)/2$$

$$\bar{x} - E(\bar{x}) = \bar{\epsilon}$$

$$\bar{y} - E(\bar{y}) = \eta,$$

we have only to show that  $\bar{\epsilon}, \eta, \bar{\epsilon}_1 - \bar{\epsilon}_2, \eta_1 - \eta_2$  are independently distributed. We obviously have

$$\bar{\epsilon} = \bar{\epsilon}_1 + \bar{\epsilon}_2, \quad \eta = \eta_1 + \eta_2$$

It is evident that  $\bar{\epsilon}_1, \bar{\epsilon}_2, \eta_1$  and  $\eta_2$  are independently distributed. Hence,  $E[\bar{\epsilon}(\bar{\epsilon}_1 - \bar{\epsilon}_2)] = (E\bar{\epsilon}_1^2 - E\bar{\epsilon}_2^2)/2 = 0$  and also  $E[\eta(\eta_1 - \eta_2)] = (E\eta_1^2 - E\eta_2^2)/2 = 0$ . Since  $\bar{\epsilon}_1 - \bar{\epsilon}_2, \eta_1 - \eta_2$ , and  $\bar{\epsilon}$  and  $\eta$  are normally distributed, the independence of this set of variables is proved, and therefore also (21) and (25) are independently distributed. It is obvious that the expression (20) is distributed independently of (21) and (25). From this it follows that

$$(26) \quad \frac{N-2}{2} \cdot \frac{N[a_1^2(a-\alpha)^2 + (\bar{y} - \alpha\bar{x} - \beta)^2]}{(N-2)s^2} \\ = \frac{(N-2)[a_1^2(a-\alpha)^2 + (\bar{y} - \alpha\bar{x} - \beta)^2]}{2[(s_y')^2 + \alpha^2(s_x')^2 - 2\alpha s_{xy}]}$$

has the  $F$ -distribution (analysis of variance distribution) with 2 and  $N-2$  degrees of freedom. The  $F$ -distribution is tabulated in Snedecor's book: *Calcu-*

lation and Interpretation of Analysis of Variance, Collegiate Press, Ames, Iowa, 1934. The distribution of  $\frac{1}{2} \log F = z$  is tabulated in R. A. Fisher's book: *Statistical Methods for Research Workers*, London, 1936. Denote by  $F_0$  the critical value of  $F$  corresponding to the chosen probability level  $P$ . Then the confidence region  $R$  is the set of points  $(\alpha, \beta)$  which satisfy the inequality

$$(27) \quad \frac{N-2}{2} \cdot \frac{a_1^2(a-\alpha)^2 + (\bar{y} - \alpha\bar{x} - \beta)^2}{(s'_y)^2 + \alpha^2(s'_x)^2 - 2\alpha s'_{xy}} < F_0.$$

The boundary of the region is given by the equation

$$(28) \quad a_1^2(a-\alpha)^2 + (\bar{y} - \alpha\bar{x} - \beta)^2 = \frac{2F_0}{N-2} [(s'_y)^2 + \alpha^2(s'_x)^2 - 2\alpha s'_{xy}].$$

This is the equation of an ellipse. Hence the region  $R$  is the interior of the ellipse defined by the equation (28). If Assumption V holds, the length of the axes of the ellipse are of the order  $1/\sqrt{N}$ , hence with increasing  $N$  the ellipse reduces to a point.

**7. The Grouping of the Observations.** We have divided the observations in two equal groups  $G_1$  and  $G_2$ ,  $G_1$  containing the first half  $(x_1, y_1), \dots, (x_m, y_m)$  and  $G_2$  the second half  $(x_{m+1}, y_{m+1}), \dots, (x_N, y_N)$  of the observations. All the formulas and statements of the previous sections remain exactly valid for any arbitrary subdivision of the observations in two equal groups, provided that the subdivision is defined independently of the errors  $\epsilon_1, \dots, \epsilon_N$ ;  $\eta_1, \dots, \eta_N$ . The question of which is the most advantageous grouping arises, i.e. for which grouping will  $a$  be the most efficient estimate of  $\alpha$  (will lead to the shortest confidence interval for  $\alpha$ ). It is easy to see that the greater  $|a_1|$  the more efficient is the estimate  $a$  of  $\alpha$ . The expression  $|a_1|$  becomes a maximum if we order the observations such that  $x_1 \leq x_2 \leq \dots \leq x_N$ . That is to say  $|a_1|$  becomes a maximum if we group the observations according to the following:

**RULE I.** The point  $(x_i, y_i)$  belongs to the group  $G_1$  if the number of elements  $x_j$  ( $j \neq i$ ) of the series  $x_1, \dots, x_N$  for which  $x_j \leq x_i$  is less than  $m = N/2$ . The point  $(x_i, y_i)$  belongs to  $G_2$  if the number of elements  $x_j$  ( $j \neq i$ ) for which  $x_j \leq x_i$  is greater than or equal to  $m$ .

This grouping, however, depends on the observed values  $x_1, \dots, x_N$  and is therefore in general not entirely independent of the errors  $\epsilon_1, \dots, \epsilon_N$ . Let us now consider the grouping according to the following:

**RULE II.** The point  $(x_i, y_i)$  belongs to the group  $G_1$  if the number of elements  $X_j$  of the series  $X_1, \dots, X_N$  for which  $X_j \leq X_i$  ( $j \neq i$ ) is less than  $m$ . The point  $(x_i, y_i)$  belongs to  $G_2$  if the number of elements  $X_j$  for which  $X_j \leq X_i$  ( $j \neq i$ ) is equal to or greater than  $m$ .

The grouping according to Rule II is entirely independent of the errors  $\epsilon_1, \dots, \epsilon_N; \eta_1, \dots, \eta_N$ . It is identical with the grouping according to Rule I in the following case: Denote by  $x$  the median of  $x_1, \dots, x_N$ ; assume that  $\epsilon$  can take values only within the finite interval  $[-c, +c]$  and that all the values  $x_1, \dots, x_N$  fall outside the interval  $[x - c, x + c]$ . It is easy to see that in this case  $x_i \leq x$  ( $i = 1, \dots, N$ ) holds if and only if  $X_i \leq X$ , where  $X$  denotes the median of  $X_1, \dots, X_N$ . Hence the grouping according to Rule II is identical to that according to Rule I and therefore the grouping according to Rule I is independent of the errors  $\epsilon_1, \dots, \epsilon_N$ . In such cases we get the best estimate of  $\alpha$  by grouping the observations according to Rule I. Practically, we can use the grouping according to Rule I and regard it as independent of the errors  $\epsilon_1, \dots, \epsilon_N; \eta_1, \dots, \eta_N$  if there exists a positive value  $c$  for which the probability that  $|\epsilon| \geq c$  is negligibly small and the number of observations contained in  $[x - c, x + c]$  is also very small.

Denote by  $a'$  the value of  $a$  which we obtain by grouping the observations according to Rule I and by  $a''$  the value of  $a$  if we group the observations according to Rule II. The value  $a''$  is in general unknown, since the values  $X_1, \dots, X_N$  are unknown, except in the special case considered above, when we have  $a'' = a'$ . We will now show that an upper and a lower limit for  $a''$  can always be given. First, we have to determine a positive value  $c$  such that the probability that  $|\epsilon| \geq c$  is negligibly small. The value of  $c$  may often be determined before we make the observations having some *a priori* knowledge about the possible range of the errors. If this is not the case, we can estimate the value of  $c$  from the data. It is well known that if we have errors in both variables and fit a straight line by the method of least squares minimizing in the  $x$ -direction, the sum of the squared deviations divided by the number of degrees of freedom will overestimate  $\sigma_\epsilon^2$ . Hence, if  $\epsilon$  is normally distributed, we can consider the interval  $[-3v, 3v]$  as the possible range of  $\epsilon$ , i.e.  $c = 3v$ , where  $v^2$  denotes the sum of the squared residuals divided by the number of degrees of freedom. If the distribution of  $\epsilon$  is unknown, we shall have to take for  $c$  a somewhat larger value, for instance  $c = 5v$ . After having determined  $c$ , upper and lower limits for  $a''$  can be given as follows: we consider the system  $S$  of all possible groupings satisfying the conditions:

- (1) If  $x_i \leq x - c$  the point  $(x_i, y_i)$  belongs to the group  $G_1$ .
- (2) If  $x_i \geq x + c$  the point  $(x_i, y_i)$  belongs to the group  $G_2$ .

We calculate the value of  $a$  according to each grouping of the system  $S$  and denote the minimum of these values by  $a^*$ , and the maximum by  $a^{**}$ . Since the grouping according to Rule II is contained in the system  $S$ ,  $a^*$  is a lower and  $a^{**}$  an upper limit of  $a''$ .

Let  $g$  be a grouping contained in  $S$  and denote by  $I_g$  the confidence interval for  $\alpha$  which we obtain from formula (23) using the grouping  $g$ . Denote further by  $I$  the smallest interval which contains the intervals  $I_g$  for all elements  $g$  of  $S$ . Then  $I$  contains also the confidence interval corresponding to the grouping according to Rule II. If we denote by  $P$  the chosen probability level (say

$P = .95$ ), then we can say: If we were to draw a sample consisting of  $N$  pairs of observations  $(x_1, y_1), \dots, (x_N, y_N)$ , the probability is greater than or equal to  $P$  that we shall obtain a system of observations such that the interval  $I$  will include the true slope  $\alpha$ .

The computing work for the determination of  $I$  may be considerable if the number of observations within the interval  $[x - c, x + c]$  is not small. We can get a good approximation to  $I$  by less computation work as follows: First we calculate the slope  $a'$  using the grouping according to Rule I and determine the confidence interval  $[a' - \delta, a' + \Delta]$  according to formula (23). Denote by  $a(g)$  the value of the slope, i.e. the value of  $\frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2}$ , corresponding to a grouping  $g$  of the system  $S$ , and by  $[a(g) - \delta_g, a(g) + \Delta_g]$  the corresponding confidence interval calculated from (23). Neglecting the differences  $(\delta_g - \delta)$  and  $(\Delta_g - \Delta)$ , we obtain for  $I$  the interval  $[a^* - \delta, a^{**} + \Delta]$ .

If the difference  $a^{**} - a^*$  is small, we can consider  $I = [a^* - \delta, a^{**} + \Delta]$  as the correct confidence interval of  $\alpha$  corresponding to the chosen probability level  $P$ . If, however,  $a^{**} - a^*$  is large, the interval  $I$  is unnecessarily large. In such cases we may get a much shorter confidence interval by using some other grouping defined independently of the errors  $\epsilon_1, \dots, \epsilon_N; \eta_1, \dots, \eta_N$ . For instance if we see that the values  $x_1, \dots, x_N$  considered in the order as they have been observed, show a monotonically increasing (or decreasing) tendency, we shall define the group  $G_1$  as the first half, and the group  $G_2$  as the second half of the observations. Though we decide to make this grouping after having observed that the values  $x_1, \dots, x_N$  show a clear trend, the grouping can be considered as independent of the errors  $\epsilon_1, \dots, \epsilon_N$ . In fact, if the range of the error  $\epsilon$  is small in comparison to the true part  $X$ , the trend tendency of the value  $x_1, \dots, x_N$  will not be affected by the size of the errors  $\epsilon_1, \dots, \epsilon_N$ . We may use for the grouping also any other property of the data which is independent of the errors.

The results of the preceding considerations can be summarized as follows:

We use first the grouping according to Rule I, calculate the slope  $a' = \frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2}$  and the corresponding confidence interval  $[a' - \delta, a' + \Delta]$  (formula (23)). This confidence interval cannot be considered as exact since the grouping according to Rule I is not completely independent of the errors. In order to take account of this fact, we calculate  $a^*$  and  $a^{**}$ . If  $a^{**} - a^*$  is small, we consider  $I = [a^* - \delta, a^{**} + \Delta]$  with practical approximation as the correct confidence interval. If, however,  $a^{**} - a^*$  is large, the interval  $I$  is unnecessarily large. We can only say that  $I$  is a confidence interval corresponding to a probability level greater than or equal to the chosen one. In such cases we should try to use some other grouping defined independently of the errors, which eventually will lead to a considerably shorter confidence interval.

Analogous considerations hold regarding the joint confidence region for  $\alpha$  and  $\beta$ . We use the grouping according to Rule I and calculate from (27) the

corresponding confidence region  $R$ . If  $|a^{**} - a^*|$  and  $|b^{**} - b^*|$  are small ( $b^* = \bar{y} - a^*\bar{x}$  and  $b^{**} = \bar{y} - a^{**}\bar{x}$ ) we enlarge  $R$  to a region  $\bar{R}$  corresponding to the fact that  $a$  and  $b$  may take any values within the intervals  $[a^{**}, a^*]$  and  $[b^{**}, b^*]$  respectively. The region  $\bar{R}$  can be considered with practical approximation as the correct confidence region. If  $|a^{**} - a^*|$  or  $|b^{**} - b^*|$  is large, we may try some other grouping defined independently of the errors, which may lead to a smaller confidence region. In any case  $\bar{R}$  represents a confidence region corresponding to a probability level greater than or equal to the chosen one.

**8. Some Remarks on the Consistency of the Estimates of  $\alpha, \beta, \sigma, \sigma_y$ .** We have shown in section 3 that the given estimates of  $\alpha, \beta, \sigma$ , and  $\sigma_y$  are consistent if condition V is satisfied.

If the values  $x_1, \dots, x_N$  are not obtained by random sampling, it will in general be possible to define a grouping which is independent of the errors and for which condition V is satisfied. We can sometimes arrange the experiments such that no values of the series  $x_1, \dots, x_N$  should be within the interval  $[x - c, x + c]$  where  $x$  denotes the median of  $x_1, \dots, x_N$  and  $c$  the range of the error  $\epsilon$ . In such cases, as we saw, the grouping according to Rule I is independent of the errors. Condition V is certainly satisfied if we group the data according to Rule I.

Let us now consider the case that  $X_1, \dots, X_N$  are random variables independently distributed, each having the same distribution. Denote by  $X$  a random variable having the same probability distribution as possessed by each of the random variables  $X_1, \dots, X_N$ . Assuming that  $X$  has a finite second moment, the expression in condition V will approach zero stochastically with  $N \rightarrow \infty$  for any grouping defined independently of the values  $X_1, \dots, X_N$ . It is possible, however, to define a grouping independent of the errors (but not independent of  $X_1, \dots, X_N$ ) for which the expression in V does not approach zero, provided that  $X$  has the following property: There exists a real value  $\lambda$  such that the probability that  $X$  will lie within the interval  $[\lambda - c, \lambda + c]$  ( $c$  denotes the range of the error  $\epsilon$ ) is zero, the probability that  $X > \lambda + c$  is positive, and the probability that  $X < \lambda - c$  is positive. The grouping can be defined, for instance, as follows:

The  $i$ -th observation  $(x_i, y_i)$  belongs to the group  $G_1$  if  $x_i \leq \lambda$  and to  $G_2$  if  $x_i > \lambda$ . We continue the grouping according to this rule up to a value  $i$  for which one of the groups  $G_1, G_2$  contains already  $N/2$  elements. All further observations belong to the other group.

It is easy to see that the probability is equal to 1 that the relation  $x_i \leq \lambda$  is equivalent to the relation  $X_i < \lambda - c$  and the relation  $x_i > \lambda$  is equivalent to the relation  $X_i > \lambda + c$ . Hence this grouping is independent of the errors. Since for this grouping condition V is satisfied, our statement is proved.

If  $X$  has not the property described above, it may happen that for every grouping defined independently of the errors, the expression in condition V con-



verges always to zero stochastically. Such a case arises for instance if  $X$ ,  $\epsilon$  and  $\eta$  are normally distributed.<sup>14</sup> It can be shown that in this case no consistent estimates of the parameters  $\alpha$  and  $\beta$  can be given, unless we have some additional information not contained in the data (for instance we know *a priori* the ratio  $\sigma_\epsilon/\sigma_\eta$ ).

**9. Structural Relationship and Prediction.**<sup>15</sup> The problem discussed in this paper was the question as to how to estimate the relationship between the true parts  $X$  and  $Y$ . We shall call the relationship between the true parts the structural relationship. The problem of finding the structural relationship must not be confused with the problem of prediction of one variable by means of the other. The problem of prediction can be formulated as follows: We have observed  $N$  pairs of values  $(x_1, y_1), \dots, (x_N, y_N)$ . A new observation on  $x$  is given and we have to estimate the corresponding value of  $y$  by means of our previous observations  $(x_1, y_1), \dots, (x_N, y_N)$ . One might think that if we have estimated the structural relationship between  $X$  and  $Y$ , we may estimate  $y$  by the same relationship. That is to say, if the estimated structural relationship is given by  $Y = aX + b$ , we may estimate  $y$  from  $x$  by the same formula:  $y = ax + b$ . This procedure may lead, however, to a biased estimate of  $y$ . This is, for instance, the case if  $X$ ,  $\epsilon$  and  $\eta$  are normally distributed. It can easily be shown in this case that for any given  $x$  the conditional expectation of  $y$  is a linear function of  $x$ , that the slope of this function is different from the slope of the structural relationship, and that among all unbiased estimates of  $y$  which are linear functions of  $x$ , the estimate obtained by the method of least squares has the smallest variance. Hence in this case we have to use the least square estimate for purposes of prediction. Even if we would know exactly the structural relationship  $Y = \alpha X + \beta$ , we would get a biased estimate of  $y$  by putting  $y = \alpha x + \beta$ .

Let us consider now the following example:  $X$  is a random variable having a rectangular distribution with the range  $[0, 1]$ . The random variable  $\epsilon$  has a rectangular distribution with the range  $[-0.1, +0.1]$ . For any given  $x$  let us denote the conditional expectation of  $y$  by  $E(y | x)$  and the conditional expectation of  $X$  by  $E(X | x)$ . Then we obviously have

$$E(y | x) = \alpha E(X | x) + \beta.$$

Now let us calculate  $E(X | x)$ . It is obvious that the joint distribution of  $X$  and  $\epsilon$  is given by the density function:

$$5 \, dX \, d\epsilon,$$

<sup>14</sup> I wish to thank Professor Hotelling for drawing my attention to this case.

<sup>15</sup> I should like to express my thanks to Professor Hotelling for many interesting suggestions and remarks on this subject.

where  $X$  can take any value within the interval  $[0, 1]$  and  $\epsilon$  can take any value within  $[-0.1, +0.1]$ . From this we obtain easily that the joint distribution of  $x$  and  $X$  is given by the density function

$$5 \, dx \, dX,$$

where  $x$  can take any value within the interval  $[-0.1, 1.1]$  and  $X$  can take any value lying in both intervals  $[0, 1]$  and  $[x - 0.1, x + 0.1]$  simultaneously. Denote by  $I_x$  the common part of these two intervals. Then for any fixed  $x$  the relative distribution of  $X$  is given by the probability density

$$\frac{dX}{\int_{I_x} dX}.$$

Hence, we have

$$E(X | x) = \frac{\int_{I_x} X \, dX}{\int_{I_x} dX}$$

We have to consider 3 cases:

$$(1) \quad 0.1 \leq x \leq 0.9.$$

In this case  $I_x = [x - 0.1, x + 0.1]$  and

$$E(X | x) = \frac{\int_{x-0.1}^{x+0.1} X \, dX}{\int_{x-0.1}^{x+0.1} dX} = x.$$

$$(2) \quad -0.1 < x \leq 0.1. \quad \text{Then } I_x = [0, x + 0.1] \text{ and}$$

$$E(X | x) = \frac{\int_0^{x+0.1} X \, dX}{\int_0^{x+0.1} dX} = .5x + .05.$$

$$(3) \quad 0.9 \leq x < 1.1. \quad \text{Then } I_x = [x - 0.1, 1] \text{ and}$$

$$E(X | x) = \frac{\int_{x-0.1}' X \, dX}{\int_{x-0.1}' dX} = .5x + .45.$$

Since

$$E(y | x) = \alpha E(X | x) + \beta,$$

we see that the structural relationship gives an unbiased prediction of  $y$  from  $x$  if  $0.1 \leq x \leq 0.9$ , but not in the other cases.

The problem of cases for which the structural relationship is appropriate also for purposes of prediction, needs further investigation. I should like to mention a class of cases where the structural relationship has to be used also for prediction. Assume that we have observed  $N$  values  $(x_1, y_1), \dots, (x_N, y_N)$  of the variables  $x$  and  $y$  for which the conditions I-IV of section 2 hold. Then we make a new observation on  $x$  obtaining the value  $x'$ . We assume that the last observation on  $x$  has been made under changed conditions such that we are sure that  $x'$  does not contain error, i.e.  $x'$  is equal to the true part  $X'$ . Such a situation may arise for instance if the error  $\epsilon$  is due to errors of measurement and the last observation has been made with an instrument of great precision for which the error of measurement can be neglected. In such cases the prediction of the corresponding  $y'$  has to be made by means of the estimated structural relationship, i.e. we have to put  $y' = ax' + b$ .

The knowledge of the structural relationship is essential for constructing any theory in the empirical sciences. The laws of the empirical sciences mostly express relationships among a limited number of variables which would prevail exactly if the disturbing influence of a great number of other variables could be eliminated. In our experiments we never succeed in eliminating completely these disturbances. Hence in deducing laws from observations, we have the task of estimating structural relationships.

COLUMBIA UNIVERSITY,  
NEW YORK, N. Y.

# A METHOD FOR MINIMIZING THE SUM OF ABSOLUTE VALUES OF DEVIATIONS

BY ROBERT R. SINGLETON

**1. Introduction.** In the *Philosophical Magazine*, 7th series, May 1930, E. C. Rhodes described a method of computation for the estimation of parameters by minimizing the sum of absolute values of deviations. His is an iterative and recursive method, in the following sense. There is a direct method for minimization with one parameter. Assuming a method for minimization with  $n - 1$  parameters, Rhodes imposes a relation between the  $n$  parameters (in an  $n$ -parameter problem) and finds a restricted minimum by the method for  $n - 1$  parameters. In this sense his method is recursive. He then repeats the process, by imposing on the  $n$  parameters a new relation determined by the restricted minimum. In this sense his method is iterative. The process is finite, ending when a restricted minimum immediately succeeds itself, indicating a true minimum.

Rhodes' paper presents the method without proof. The purpose of the present paper is to analyze the situation in detail sufficient to indicate proofs for various methods, and to present a new method which reduces the labor of solution by eliminating the recursive feature. The iterative approach is retained. The solution of Rhodes' illustrative problem will be given for comparison between the two methods.

The paper uses geometric terminology and develops to quite an extent the geometry of a surface representing the summed absolute deviations. This seems the clearest means of presenting the relationships. Further analysis of the properties of this surface should lead to an even more direct method for attaining the minimum than the one here presented.

In the writing of the paper, no attention has been given to sets of observations or equations among which a linear dependence may exist. In practice, such a situation almost never occurs. If the need arises, the adjustments which must be made to take care of dependence are in each case fairly obvious.

**2. Geometric Analogue of Summed Absolute Deviations.** Let  $n$  observations on  $\nu + 1$  variates be represented by  $x_\alpha^i, y^i$  where  $i = 1, \dots, n; \alpha = 1, \dots, \nu$ . Unless otherwise noted, latin indices have range 1 to  $n$ , greek indices, 1 to  $\nu$ . The summation convention of tensor analysis is used.

The variates are to be statistically related by the linear function<sup>1</sup>

$$\hat{y}^i = x_\alpha^i u^\alpha,$$

---

<sup>1</sup> This includes the linear function with a constant, since a variate  $x^i = 1$  may be used.

$y^i$  being an estimate of  $y^i$ .  $u^\alpha$  are to be determined so that  $v = \sum_i |y^i - y^i|$  is a minimum. Set

$$(1) \quad v^i = x_a^i u^\alpha - y^i$$

and determine functions  $e^i(u^\alpha)$  so that  $e^i v^i \geq 0$ , and  $|e^i| = 1$ . It is immaterial that  $e^i$  is not uniquely determined when  $u^\alpha$  satisfies  $v^i = 0$ . Then  $v = \sum_i e^i v^i$  is to be minimized. Using (1),

$$(2) \quad v = x_a u^\alpha - y$$

where

$$x_a = \sum_i e^i x_a^i, \quad y = \sum_i e^i y^i.$$

Consider a Euclidean  $(\nu + 1)$ -space,  $E_{\nu+1}$ , with coordinates  $u^1, \dots, u^\nu, v$ . The coordinate hyperplane perpendicular to the  $v$ -axis will be called  $E_\nu$ . In  $E_{\nu+1}$  each of equations (1) for a particular  $i$  represents a  $\nu$ -plane which intersects  $E_\nu$  in a  $(\nu - 1)$ -plane when  $v^i = 0$ . Each of the equations

$$(3) \quad v^i = e^i(x_a^i u^\alpha - y^i)$$

represents two half-planes which touch  $E_\nu$  and each other along the  $(\nu - 1)$ -plane given in  $E_\nu$  by the equation

$$(4) \quad x_a^i u^\alpha - y^i = 0.$$

The functions on the right-hand side of (3) are thus continuous everywhere, and linear in any neighborhood of  $E_\nu$ , none of whose points satisfies (4). Since a sum of functions continuous and linear in a neighborhood is also continuous and linear in that neighborhood, it follows that the function on the right in (2) is continuous for all  $u$ , and linear for every neighborhood of  $E_\nu$  containing no points which satisfy (4) for any  $i$ . Hence

**OBSERVATION I:** *The surface ( $S$ ) given in  $E_{\nu+1}$  by (2) consists of portions of  $\nu$ -planes joined together. The projection of these joins on  $E_\nu$  forms a network of  $(\nu - 1)$ -planes determined in  $E_\nu$  by equations (4).*

**3. Existence of a Minimum.** Define a "bend of degree  $r$  on  $S$ " to be the locus of all points on  $S$  whose  $u$ -coordinates satisfy a set of  $r$  independent equations of (4). To each set of  $r$  independent equations corresponds a unique bend of degree  $r$ .

If a linear relation  $u^\alpha = a_\sigma^\alpha \lambda^\sigma + b^\alpha$ ,  $\sigma = 1, \dots, \mu < \nu$ ,  $\text{rank}(a_\sigma^\alpha) = \mu$ , is imposed on  $u^\alpha$ , all the preceding development, reduced in dimension, applies to the new variates  $x_a^i a_\sigma^i$ ,  $y^i - x_a^i b^\alpha$ .

**OBSERVATION II:** *A section of  $S$  by a plane of any dimension  $d < \nu$  has all the properties of an  $S$ -surface of dimension  $d$ .*

Since any set of consistent equations selected from (4) determines such a linear relation for  $u^\alpha$ , the application of Observation I to any of the bends of  $S$  shows that each  $r$ -bend consists of linear elements of dimension  $\nu - r$ , joined

at points which lie on linear elements of lesser dimension. Thus  $S$  is a polyhedron. Its faces we term complexes of dimension  $\nu$ ,  $C_\nu$ , and the linear elements of its edges which lie wholly in bends of degree  $r$ , but not of degree  $r + 1$  are complexes  $C_{\nu-r}$  of dimension  $\nu - r$ . The boundary of any  $C_\alpha$ ,  $\alpha > 0$ , consists of complexes of lesser dimension. The term complex is not restricted to either open or closed complexes.

Since the function  $v(u^\alpha)$  of (2) is non-negative, it possesses a greatest lower bound (g.l.b.)  $g$ . Since for some number  $h > g$ , there exists an  $N$  such that for all  $|u^\alpha| > N$ ,  $v(u^\alpha) > h$ , it follows that for some closed neighborhood of  $E$ , the g.l.b. of  $v$  is  $g$ . Since  $v$  is continuous everywhere it attains its g.l.b., and so  $S$  has minimum points. Since the minimum of any complex not parallel to  $E$ , lies on its boundary, and the boundary consists of complexes, it follows that the minimum points of  $S$  consist of  $C_0$ 's and/or entire complexes of dimension  $> 0$  which are parallel to  $E$ . The next section will show that  $S$  has a unique minimum complex (including of course its boundary complexes) and furthermore is cup-shaped.

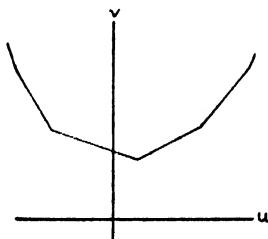


FIG. 1

**4. Convexity Property; Uniqueness of the Minimum.** Consider  $\nu = 1$  in the preceding treatment (and for convenience not written).  $S$  looks generally like Fig. 1. The slope changes only where an equation of (4) has a root. Suppose the point is  $u_0$ , and  $x^1 u_0 - y^1 = 0$ . From (3), since  $v^1 \geq 0$ , it follows that  $e^1 x^1 < 0$  for  $u < u_0$ ,  $e^1 x^1 > 0$  for  $u > u_0$ . Since in (2)  $x = \sum e^i x^i$ , and since for  $h$  sufficiently small and  $u_0 - h < u < u_0 + h$  the only  $e$  to change value<sup>2</sup> is  $e^1$ , we have that

$$x(u_1) + 2 |e^1 x^1| = x(u_2)$$

where

$$u_0 - h < u_1 < u_0 < u_2 < u_0 + h.$$

Hence the slope is a monotonic increasing step function. Since for  $u$  sufficiently small all  $e^i x^i < 0$ , and for  $u$  sufficiently large all  $e^i x^i > 0$ , at some intermediate point or points either the slope is zero or it changes from negative to

<sup>2</sup> The  $e$ 's corresponding to equations proportional to equation (1) also change value at  $x_0$ . This does not destroy the argument.

positive without becoming zero. In the first case a single closed  $C_1$  is the minimum complex; in the second, a  $C_0$ . In either case the curve given by (2) when  $\nu = 1$  is concave upward and has just one minimum complex, except for complexes of lesser dimension constituting the boundary of this complex. An obvious consequence is

**LEMMA I.** *The set of points  $u$  for which  $\nu$  is less than some number  $N$  form a convex point set.*

This result is easily extended to the general dimension  $\nu$ . If for any two points  $u_1, u_2$  of  $E_r$ ,  $\nu(u_1) < N$  and  $\nu(u_2) < N$ , the plane in  $E_{r+1}$  given by  $u^\alpha = u_1^\alpha + \lambda(u_2^\alpha - u_1^\alpha)$  makes a one-dimensional section of  $S$ . By Observation II, the points  $u$  lying on the projection of this section on  $E_r$  have the property of Lemma I and of course lie on the straight line joining  $u_1$  and  $u_2$ . This is the property required for a convex point set. Hence

**THEOREM I.** *The set of points  $u^\alpha$  of  $E_r$  for which  $\nu(u^\alpha)$  as given by (2) is less than a fixed quantity form a convex point set.*

From this it follows immediately that there is a unique minimum complex. It is appropriate here to point out that no two complexes can be contained in a single plane of the same dimension. This follows from the equation giving monotonicity of slope in one dimension, and Observation II.

**5. Gradient Directions.** From here on the treatment will be of  $\nu$  as a function defined on  $E_r$ , and the equations will represent objects in  $E_r$ , unless otherwise stated. Complex and Bend also will refer to the projections on  $E_r$  of the complexes and bends of  $S$ . For a single-valued function defined on  $E_r$  the gradient at a point is the projection of a normal to the surface representing the function in  $E_{r+1}$ . If the function is defined only over a subspace of  $E_r$  possessing derivatives, the gradient will be required also to be tangent to the subspace. This is sufficient to determine a unique direction, and preserves the property that for an infinitesimal displacement in any direction the value of the function decreases most rapidly in the direction of the gradient. Here gradient is taken negative to its usual sense.

A point  $u$  lying on a  $C_r$  but not on a  $C_{r-1}$  will have a gradient in  $C_r$  and also in each higher-dimensional complex on whose boundary  $C_r$  lies. If the gradient for  $u$  as a point of  $C_{r+k}$  points into  $C_{r+k}$  (remembering that  $u$  lies on the boundary) this will be called a usable gradient. In the case of the greatest  $k$  for which there exists a usable gradient, there exists but one  $C_{r+k}$  providing such a gradient, and that gradient is the "best" gradient; that is, of all directions in  $E_r$  it provides the direction of most rapid decrease of the function  $\nu$ . This follows from Theorem I. Furthermore, all complexes of lesser dimension providing usable gradients lie on the boundary of this  $C_{r+k}$ . In fact

**THEOREM II.** *If for a point  $u$  on  $C_r$ , two complexes  $C_s$  and  $C'_s$ ,  $s > r$ , lying in different bends of degree  $\nu - s$  but incident at  $C_r$ , both provide usable gradients for  $u$ , then the complex  $C_{s+1}$  on whose boundary lie both  $C_s$  and  $C'_s$  also provides a usable gradient for  $u$ .*

This follows from Theorem I. Select  $u_1$  on the gradient in  $C_*$ ,  $u_2$  on the gradient in  $C'_*$ , for which  $v(u_1) = v(u_2)$ . The join of  $u_1$  and  $u_2$  lies in  $C_{*+1}$ , and for some point,  $u_3$  on this join,  $v(u_3)$  is less than  $v(u_1) = v(u_2)$ . Also, the distance  $\overline{u_1 u_3}$  is less than at least one of  $\overline{u_1 u_2}$ ,  $\overline{u_2 u_3}$ . Hence  $C_{*+1}$  must contain a usable gradient.

**6. Selection of Best Gradient at Bends.** The direction of the gradient for a point  $u_0$  considered as lying on a  $C_*$  is given by

$$(5) \quad g^a = -x_a(u_0) = -\sum_i e^i(u_0) x_a^i.$$

If  $u_0$  lies in the interior of a face, this is unique. If  $u_0$  lies in a bend, so that some  $e^i$  are not determined, the  $g^a$  for each face is found by selecting the indeterminate  $e^i$ 's as  $+1$  or  $-1$ , according to the face being considered.

For a point  $u_0$  considered as lying on a bend of degree  $r$ , given by  $r$  independent equations of (4):

$$(6) \quad x_a^\lambda u^a - y^\lambda = 0, \quad (\lambda = 1, \dots, r),$$

the gradient for a particular  $C_{*+r}$ , determined by the conditions at the beginning of section 5, is

$$(7) \quad g^a = x_a^\lambda k_\lambda - x_a$$

where  $k_\lambda$  satisfies

$$\sum_a x_a^\mu x_a^\lambda k_\lambda = \sum_a x_a^\mu x_a, \quad (\mu = 1, \dots, r)$$

and  $x_a$  is as given in (2), the choice of sign for the indeterminate  $e^\lambda$  ( $\lambda = 1, \dots, r$ ) being immaterial. They may, in fact, be taken as 0 in this instance.

For a point  $u_0^\sigma$  lying on an  $r$ -bend given by (6), to determine which complex contains the best gradient, each  $(r-1)$ -bend incident on the  $r$ -bend at  $u_0$  is tested for a usable gradient. Theorem II then determines the complex containing the best gradient.

There are  $2r$  such complexes incident at  $u_0$ , given by the  $r$  sets of equations selected from (6):

$$(8) \quad (\lambda): x_a^\sigma u^a - y^\sigma = 0 \quad (\sigma = 1, \dots, \lambda-1, \lambda+1, \dots, r) \\ (\lambda = 1, \dots, r).$$

The two complexes lying in the same  $(r-1)$ -bend have the same equations in (8), but are distinguished later by  $e^\lambda(u_0)$  for the omitted equation being taken first  $+1$ , then  $-1$ .

The gradient for the  $\lambda$ th pair of complexes is

$$g_\lambda^a = x_a^\sigma k_\sigma - x_a$$

similar to (7), but not identical. For  $e^\lambda = +1$  in determining  $x_a$ , we have  $g_{\lambda+}^a$ , and for  $e^\lambda = -1$ ,  $g_{\lambda-}^a$ . We restrict the consideration to  $e^\lambda = +1$ .



The line in the direction of greatest slope is then

$$u^\alpha = u_0^\alpha + g_{\lambda+}^\alpha t.$$

Now  $u_0$  is here considered lying on the complex given by (8 $\lambda$ ) with  $e^\lambda = +1$ . In order that  $g_{\lambda+}^\alpha$  point into this face, the deviation for the  $\lambda$ th observation must exceed 0 when  $t > 0$ ; otherwise, for a displacement in the direction of  $g_{\lambda+}^\alpha$ ,  $e^\lambda$  changes sign immediately and the course is in the other complex. This deviation is

$$v^\lambda = x_\alpha^\lambda u^\alpha - y^\lambda = x_\alpha^\lambda u_0^\alpha - y^\lambda + x_\alpha^\lambda g_{\lambda+}^\alpha t = x_\alpha^\lambda g_{\lambda+}^\alpha t.$$

Had  $g_{\lambda-}^\alpha$  been used, this deviation must be less than 0. Hence a necessary and sufficient condition that a complex given by (8) with either choice of  $e^\lambda$  possess a usable gradient is

$$(9) \quad \Phi_\lambda = e^\lambda [\Sigma_\alpha x_\alpha^\lambda x_\alpha^\sigma k_\sigma - \Sigma_\alpha x_\alpha^\lambda x_\alpha] > 0.$$

For  $r = 1$  the condition is given by (9) with the first sum merely omitted.  $\Phi_{\lambda+}$  and  $\Phi_{\lambda-}$  cannot both exceed 0.

When all sets of equations (8 $\lambda$ ) are tested by (9) the equations common to all sets possessing a usable gradient determine the complex with the best gradient, retaining the values of  $e$  for which (9) was satisfied.

**7. Property of the Minimum Point.** For a minimum point, given by (6) with  $r = \nu$ , all  $\Phi_\lambda$  must be negative. Define  $X^{\beta\gamma} = \Sigma_\alpha x_\alpha^\beta x_\alpha^\gamma$  and  $X^{\beta 0} = \Sigma_\alpha x_\alpha^\beta x_\alpha$  for convenience. Then in (9), the numbers  $k_\sigma$ ,  $-1$  are seen from their definition in (7) to be proportional to the cofactors of the  $\lambda$ th row of the matrix  $(X^{\mu\sigma}, X^{\mu 0})$ ,  $\mu$  having the same range as  $\lambda$ . Thus  $\Phi_{\lambda+} = c \text{Det}(X^{\mu\sigma}, X_{+}^{\mu 0})$ , and  $\Phi_{\lambda-} = -c \text{Det}(X^{\mu\sigma}, X_{-}^{\mu 0})$ , where in the first case  $X^{\mu 0}$  is determined with  $e^\lambda = +1$ , in the second with  $e^\lambda = -1$ . The factor of proportionality,  $c$ , must be the same since  $X^{\mu\sigma}$  is unaffected by change of  $e^\lambda$ . Now let  $X^\mu = \Sigma_\alpha x_\alpha^\mu x_\alpha^*$  where  $x_\alpha^* = \Sigma_k e^k x_\alpha^k$ , the range of  $k$  omitting the range of  $\lambda$ . Then

$$\Phi_{\lambda+} = c [\text{Det}(X^{\mu\sigma}, X^\mu) + \text{Det}(X^{\mu\sigma}, X^{\mu\lambda})]$$

and

$$\Phi_{\lambda-} = -c [\text{Det}(X^{\mu\sigma}, X^\mu) - \text{Det}(X^{\mu\sigma}, X^{\mu\lambda})].$$

Hence

$$\Phi_{\lambda+}\Phi_{\lambda-} = -c^2 \{[\text{Det}(X^{\mu\sigma}, X^\mu)]^2 - [\text{Det}(X^{\mu\sigma}, X^{\mu\lambda})]^2\}.$$

Now let  $A$  represent the square matrix  $(x_\alpha^\sigma)$ ,  $\alpha$  giving the rows and  $\lambda$  the columns. Let  $B_\lambda$  represent the matrix formed from  $A$  by replacing the  $\lambda$ th column by  $x_\alpha^*$ . Then

$$\begin{aligned} \Phi_{\lambda+}\Phi_{\lambda-} &= -c^2 [\text{Det}^2(A'B_\lambda) - \text{Det}^2(A'A)] \\ &= -c^2 \text{Det}^2 A (\text{Det}^2 B_\lambda - \text{Det}^2 A) \end{aligned}$$

and this will have the same sign as

$$\Psi_{\lambda} = |\text{Det}(A)| - |\text{Det}(B_{\lambda})|.$$

Since  $\Phi_{\lambda+}$  and  $\Phi_{\lambda-}$  are never both positive, and at the minimum are both negative for all  $\lambda$ , at the minimum all  $\Psi_{\lambda} > 0$ . To determine all  $\Psi_{\lambda}$  together, let, in matrix notation,  $z' = (z_1, \dots, z_r)$  and  $x^{*'} = (x_1^*, \dots, x_r^*)$  where  $x_a^*$  were defined previously. Determine  $z$  as the solution of  $Az = x^*$ . Then  $|\text{Det}(B_{\lambda})|$  are equal to  $|z_{\lambda}| |\text{Det}(A)|$ . Hence a necessary and sufficient condition that  $\Psi_{\lambda} > 0$  for all  $\lambda$  is that all  $|z_{\lambda}|$  be less than one. Hence

**THEOREM III:** *If a zero-complex is given by a set of equations whose matrix is  $M$ , a necessary and sufficient condition that the complex be a unique minimum is that the solutions of  $M'z = x^*$  be all less than one in absolute value. If  $k$  of the solutions are equal to one in absolute value, and the rest are less than one, the minimum is a complex of dimension  $k$  with the zero-complex as one of its corners.*

The last statement follows since if one solution is 1 in absolute value, a corresponding  $\Phi_{\lambda} = 0$ , and hence no gradient, usable or not, exists. Thus the corresponding complex is parallel to  $E_r$ .

**8. Minimization for One Dimension.** A method for minimization of (2) when there is just one parameter evolves from the monotonicity of slope in that case. Suppose the variates are  $w^i$  and  $z^i$ , and (1) is

$$(10) \quad v^i = w^i t - z^i.$$

Suppose the variates are arranged in order of  $z^i/w^i$ , starting with the smallest. The slope of the  $r$ th segment (Fig. 1) from the left is

$$\sum_{i=1}^r |w^i| - \sum_{i=r+1}^n |w^i|.$$

The minimum occurs when the slope is 0 or changes from negative to positive; that is, when the first sum equals or exceeds the second; or when the first sum equals or exceeds half the total. This is a standard computation. If the change takes place when  $r = k$ , then  $t = z^k/w^k$  is the value of  $t$  giving the minimum.

**9. Minimization Procedure for  $\nu + 1$  Dimensions.** For any continuous function with unique minimum and having the property of Theorem I, the following holds. Let  $u_0$  be any point of  $E_r$ . Let  $u_{i+1} = u_i + \lambda_i t_i$ , where  $\lambda_i$  is any direction chosen at random and  $t_i$  is the value of  $t$  for which the function attains a minimum on the curve  $u = u_i + \lambda_i t$ . Then the probability is one that  $\lim_{i \rightarrow \infty} u_i = u_1$ , where  $u_1$  is a minimum point for the function. If  $\lambda_i$  is taken always as the gradient of  $u_i$ , such a procedure is called the "method of steepest descent" for approaching the minimum point.

Usually the limit is never attained. In this case, however, the minimum is

TABLE I  
*Method of Steepest Descent Applied*

Data					First Restricted Minimum					Second Restricted Minimum					Third (absolute) Minimum					for test
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
$i$	$x_1^i$	$x_2^i$	$x_3^i$	$y^i$	$z_0^i$	$e^i(u_0)$	$w_0^i$	$z_0^i$	rank of (9)	$z_1^i$	$e_1^i$	$w_1^i$	$z_1^i$	rank of (14)	$z_2^i$	$e_2^i$	$w_2^i$	$z_2^i$	rank of (19)	$e_3^i$
1	1	-8	64	28	-78	-	-5165	.015	16	-47.044	-	-1149.570	-.0409	6	-92.773	-	-285.180	.3253	15	-
2	1	-7	49	-39	14	+	-3917	-.004	8	37.476	+	-942.102	.0398	12	0	0	0	.0933	13	+
3	1	-6	36	-40	36	+	-2841	-.013	6	53.027	+	-753.324	.0704	14	23.060	+	247.156	.0933	13	+
4	1	-5	25	-1	14	+	-1937	-.007	7	25.609	+	-583.236	.0439	13	2.408	+	456.288	.0053	8	+
5	1	-4	16	23	3	+	-1205	-.002	9	10.222	+	-431.838	.0237	11	-6.956	-	627.396	-.0111	2	-
6	1	-3	9	34	1	+	-645	-.0015	10	4.866	+	-299.130	.0163	10	-7.033	-	760.480	-.0092	3	-
7	1	-2	4	25	15	+	-257	-.058	2	16.540	+	-185.112	.0894	15	9.177	+	855.540	.0107	9	+
8	1	-1	1	40	1	+	-41	-.024	3	1.246	+	-89.784	.0139	8	-2.326	-	912.576	-.0025	5	-
9	1	0	0	43	-5	-	3	-1.67	1	-5.018	-	-13.146	-.3817	2	-5.541	-	931.588	-.0059	4	-
10	1	1	1	29	2	+	-125	-.016	4	2.749	+	44.802	-.0614	5	4.531	+	912.576	.0050	7	10
11	1	2	4	14	6	+	-425	-.014	5	8.547	+	84.060	-.1017	4	11.891	+	855.540	.0139	10	+
12	1	3	9	12	-7	-	-897	.008	14	-1.624	-	104.628	.0155	9	2.538	+	760.480	.0033	6	-
13	1	4	16	-16	2	+	-1541	-.001	11	11.236	+	106.506	-.1055	3	15.472	+	627.396	.0247	12	+
14	1	5	25	-14	-23	-	-2357	.010	15	-8.874	-	89.604	.0989	16	-5.306	-	456.288	-.0116	1	-
15	1	6	36	-46	-18	-	-3345	.005	12	2.048	+	54.192	-.0378	7	4.203	+	247.156	.0170	11	+
16	1	7	49	-68	-27	-	-4505	.006	13	0	0	0	0	1	0	0	0	0	11	+
17	1	8	64	-24	-106	-	-5837	.018	17	-71.017	-	-72.882	-.9744	1	-73.916	-	-285.180	.2592	14	-
$\Sigma$	17	0	408	0	-170	-	-35037			39.990	-	-4036.242			-120.570		8080.100			

attained. The minimum can be approached as closely as desired, hence a complex incident on the minimum is reached. But the convex point sets of Theorem I surrounding the minimum complex are all similar convex polyhedrons in  $E_r$ , whose corresponding faces are parallel, and the gradients at points on a bend cannot point into a higher dimensional complex on the bend. Hence the sequence of points lie on bends of successively greater degree, and must eventually attain the minimum complex.

TABLE II

*Points  $u_k$* 

$$u_{k+1}^a = u_k^a + g_k^a t_k$$

$$u_0 = (38, -5, -2)$$

$$u_1 = (37.98202, -4.74828, -1.48457)$$

$$u_2 = (37.45908, -2.07142, -1.85631)$$

$$u_3 = (32.83333, -2.07142, -1.76191)$$

TABLE III

*Computation of  $t_k = z_k/w_k$* 

$\Sigma  w_k $	in order of col.	exceeds	at $i =$	hence $t_k =$
$\Sigma  w_0 $	(10)	17521	16	.00599334
$\Sigma  w_1 $	(15)	2502	2	.0397792
$\Sigma  w_2 $	(20)	4610	10	.00496545

TABLE IV

*Gradients  $g_k^a$  for column  $(5k + 8)$* 

$k$	$g_k^1$	$g_k^2$	$g_k^3$
0	-3	42	86
1	-13146	67293	-9345
2	-931588	0	19012

The computational procedure is as follows:

1. Select a point  $u_0$ .
2. Determine the gradient  $g_0^a$  from (5).
3. Compute  $w_0^i = x_a^i g_0^a$ ,  $z_0^i = y^i - x_a^i u_0^a$ .
4. Determine  $t_0$  by the method of section 8.
5. Compute  $u_1^a = u_0^a + g_0^a t_0$ .
6. Determine the complex containing the best gradient by (9), and the gradient  $g_1^a$  by (7).

and so proceed to the minimum. This may be finally tested by Theorem III.

Step 5 is unnecessary, since the only use for  $u_1^a$  is to determine  $e^i(u_1)$ . But  $e^i(u_1) = e^i(t_0)$ , the latter referring to the computation in step 4. Also, after the first step, it is easier to compute  $z^i$  by

$$z_{k+1}^i = z_k^i - w_k^i t_k.$$

**10. Example.** The computation for (9) is not so great as it would seem, since some of the work is duplication and some must be computed anyway for the gradient. Even so, for  $r \geq 3$  it becomes, perhaps, more arduous than its contribution would seem to justify. For  $\nu \geq 4$  it is recommended that the test of (9) be omitted for points on bends of third degree or greater, and the final test of Theorem III be applied at the end of the work. If this test shows the minimum has not been reached, the complex in which lies the best gradient will be indicated at the same time.

The minimum number of steps is 0. The maximum number is tremendous but finite. The expected number is probably a little greater than  $\nu$ .

In Tables I to IV, the method is applied to the problem used by Rhodes to illustrate his method. The independent variates are shown in columns (2), (3), (4), Table I, the dependent variate in column (5). The only other original datum is the initial point, selected by guess, shown in line 1, Table II. Since slightly different formulas were used in the computation, the signs of cols. (6), (8), (11), (16), (18) are reversed, and the gradients in Table IV are multiplied by constants. As they are used only for directions, this does not matter.

PRINCETON UNIVERSITY,  
PRINCETON, N. J.

# A STUDY OF A UNIVERSE OF $n$ FINITE POPULATIONS WITH APPLICATION TO MOMENT-FUNCTION ADJUSTMENTS FOR GROUPED DATA

BY JOSEPH A. PIERCE

The object of this paper is to study the case of a universe of  $n$  finite populations, considering both the expectations of population moment-functions and the moments of sample moments, and to make applications of the results which may be of interest to mathematical statisticians. The sampling formulas which are derived reduce to the usual infinite or finite sampling formulas, under appropriate assumptions. Also a method is given whereby finite sampling formulas may be transformed into the corresponding infinite sampling formulas.

The general methods and formulas which are given in Part I for the expectations of population moment-functions are used, in Part II, to find the expectations of moments of a distribution of discrete data grouped in " $k$  groupings of  $k$ ".

## I. A STUDY OF A UNIVERSE OF $n$ FINITE POPULATIONS

Let  ${}_nU_N$  be a universe composed of the set of populations  ${}_rX$ , ( $r = 1, 2, \dots, n$ ) each population  ${}_rX$  consisting of a finite number of discrete variates  ${}_rx_i$ , ( $i = 1, 2, \dots, N$ ), ( $N > n$ ). The  $t$ th moment of  ${}_rX$  is denoted by  ${}_r\mu_t$ . The  $t$ th central moment of  ${}_rX$  is denoted by  ${}_r\bar{\mu}_t$ . The  $t$ th moment and the  $t$ th central moment of  ${}_nU_N$  are respectively denoted by  $\mu_t$  and  $\bar{\mu}_t$ . The expected value of a variable  $y$  is denoted by  $E(y)$ . We have

$$\begin{aligned} {}_r\mu_t &= E({}_rx_i^t) = \frac{1}{N} \sum_{i=1}^N {}_rx_i^t, & {}_r\bar{\mu}_t &= E({}_rx_i - {}_r\mu_1)^t = \frac{1}{N} \sum_{i=1}^N ({}_rx_i - {}_r\mu_1)^t, \\ (1.1) \quad \mu_{1:\mu_t} &= E({}_r\mu_t) = \frac{1}{n} \sum_{r=1}^n {}_r\mu_t, & \mu_{1:\bar{\mu}_t} &= E({}_r\bar{\mu}_t) = \frac{1}{n} \sum_{r=1}^n {}_r\bar{\mu}_t, \\ \mu_{s_1 s_2 \dots s_v; \mu_{t_1} \mu_{t_2} \dots \mu_{t_v}} &= E({}_r\mu_{t_1}^{s_1} {}_r\mu_{t_2}^{s_2} \dots {}_r\mu_{t_v}^{s_v}), \\ \mu_{s_1 s_2 \dots s_v; \bar{\mu}_{t_1} \bar{\mu}_{t_2} \dots \bar{\mu}_{t_v}} &= E({}_r\bar{\mu}_{t_1}^{s_1} {}_r\bar{\mu}_{t_2}^{s_2} \dots {}_r\bar{\mu}_{t_v}^{s_v}). \end{aligned}$$

We also note that  $\mu_{s_1 s_2 \dots s_v; \mu_{t_1} \mu_{t_2} \dots \mu_{t_v}}$  may be written  $\mu_{111 \dots 1; t_1 \mu_{t_1}^{s_1} t_2 \mu_{t_2}^{s_2} \dots t_v \mu_{t_v}^{s_v}}$ .

**1. The expected value of moments and central moments.** It follows easily from (1.1) that

$$(1.2) \quad \mu_{1:\mu_t} = \mu_t.$$

From the usual formula for central moments in terms of moments, we get

$$(1.3) \quad \mu_{1;\bar{\mu}_t} = \sum_{i=0}^t (-1)^i \binom{t}{i} \mu_{i1;\mu_1\mu_{t-i}}.$$

Terms of the form  $\mu_{i1;\mu_1\mu_{t-i}}$  may be evaluated by use of the well known formulas [20; p. 58] for changing from moments to central moments in the case of a multivariate distribution. Two of these formulas are given below.

$$(1.4) \quad \begin{aligned} \bar{\mu}_{11;\mu_a\mu_b} &= \mu_{11;\mu_a\mu_b} - \mu_{10;\mu_a\mu_b}\mu_{01;\mu_a\mu_b} \\ \bar{\mu}_{111;\mu_a\mu_b\mu_c} &= \mu_{111;\mu_a\mu_b\mu_c} - \mu_{110;\mu_a\mu_b\mu_c}\mu_{001;\mu_a\mu_b\mu_c} \\ &\quad - \mu_{101;\mu_a\mu_b\mu_c}\mu_{010;\mu_a\mu_b\mu_c} - \mu_{011;\mu_a\mu_b\mu_c}\mu_{100;\mu_a\mu_b\mu_c} \\ &\quad + 2\mu_{100;\mu_a\mu_b\mu_c}\mu_{010;\mu_a\mu_b\mu_c}\mu_{001;\mu_a\mu_b\mu_c}. \end{aligned}$$

We find that

$$(1.5) \quad \mu_{i1;\mu_1\mu_{t-i}} = \sum \frac{i!}{p_1!p_2!r_1!r_2!} \bar{\mu}_{p_1r_1;\mu_1\mu_{t-i}} \mu_{1;\mu_1}^{p_2} \mu_{1;\mu_{t-i}}^{r_2},$$

where  $p_1p_2$  is a two-part partition of  $i$  and  $r_1 + r_2 = 1$ .

Using (1.3) and (1.5), we get

$$(1.6) \quad \mu_{1;\bar{\mu}_2} = \bar{\mu}_2 - \bar{\mu}_{2;\mu_1}.$$

$$(1.7) \quad \mu_{1;\bar{\mu}_3} = \bar{\mu}_3 - 3\bar{\mu}_{11;\mu_1\mu_2} + 6\mu_1\bar{\mu}_{2;\mu_1} + 2\bar{\mu}_{3;\mu_1}.$$

$$(1.8) \quad \begin{aligned} \mu_{1;\bar{\mu}_4} &= \bar{\mu}_4 + 6(\bar{\mu}_2 - 2\mu_1^2) \bar{\mu}_{2;\mu_1} - 12\mu_1\bar{\mu}_{3;\mu_1} + 12\mu_1\bar{\mu}_{11;\mu_1\mu_2} \\ &\quad - 4\bar{\mu}_{11;\mu_1\mu_2} + 6\bar{\mu}_{21;\mu_1\mu_2} - 3\bar{\mu}_{4;\mu_1}. \end{aligned}$$

etc.

If the  $n$  populations are identical, it is evident from the definition of  $\mu_{1;\bar{\mu}_t}$  that, for all finite  $t$ ,

$$\mu_{1;\bar{\mu}_t} = \bar{\mu}_t.$$

**2. The expected value of Thiele seminvariants.** If the  $t$ th Thiele seminvariant is denoted by  $\lambda_t$ , then

$$(1.9) \quad \mu_{1;\lambda_t} = \sum \frac{(-1)^{s-1} t! (\rho - 1)!}{s_1! s_2! \dots s_v! (2!)^{s_2} (3!)^{s_3} \dots (v!)^{s_v}} \mu_{s_1 s_2 \dots s_v; \mu_1 \mu_2 \dots \mu_v},$$

the summation being taken for all positive integers  $s_i$  ( $i = 1, 2, \dots, v$ ), for which

$$\rho = \sum_{i=1}^v s_i, \quad t = \sum_{i=1}^v i s_i.$$

Terms of the form  $\mu_{s_1 s_2 \dots s_v; \mu_1 \mu_2 \dots \mu_v}$  are evaluated by (1.4). We have

$$(1.10) \quad \mu_{1;\lambda_2} = \lambda_2 - \bar{\mu}_{2;\mu_1}.$$

$$(1.11) \quad \mu_{1:\lambda_3} = \lambda_3 - 3\bar{\mu}_{11:\mu_1\mu_2} + 6\lambda_1\bar{\mu}_{2:\mu_1} + 2\bar{\mu}_{3:\mu_1}.$$

$$(1.12) \quad \begin{aligned} \mu_{1:\lambda_4} = & \lambda_4 + 12[\lambda_2 - 2\lambda_1^2] \bar{\mu}_{2:\mu_1} - 24\lambda_1\bar{\mu}_{3:\mu_1} + 24\lambda_1\bar{\mu}_{11:\mu_1\mu_2} \\ & - 4\bar{\mu}_{11:\mu_1\mu_3} + 12\bar{\mu}_{21:\mu_1\mu_2} - 3\bar{\mu}_{2:\mu_2} - 6\bar{\mu}_{4:\mu_1}. \end{aligned}$$

etc.

If the  $n$  populations are identical then, for all finite  $t$ ,

$$\mu_{1:\lambda_t} = \lambda_t.$$

**3. Generalized sampling.** It follows from definition that all rational isobaric moment-functions have the property that they may be expressed in terms of power sums and power product sums with certain coefficients. Of the power sums and power product sums which enter a sampling formula only the power product sums take different forms depending on the law of variate selection. Now, there are two possible courses which may be followed by one who wishes to derive sampling formulas for the case of a single population.

1. One may decide in advance on the law which he wishes to govern the selection of variates which enter the sample. Then he may apply this law in the evaluation, in terms of moments, of every power product term as it occurs in each formula which is derived.

2. One may derive the formulas for sampling under the condition that the law is unspecified, thereby obtaining formulas which are capable of being interpreted in terms of laws that are decided upon later.

We illustrate the two possible courses by considering the formula,

$$(1.13) \quad \bar{\mu}_{2:z} = \frac{r}{s} \Sigma \bar{x}^2 + \frac{2r(r-1)}{s(s-1)} \Sigma \bar{x}_i \bar{x}_j,$$

which Carver [12; p. 102] obtains for the case of finite sampling without replacements. Here  $r$  = the number in the sample,  $s$  = the number in the parent population and  $z_i$  = the algebraic sum of the variates of  $i$ th sample. Later, by evaluating  $\Sigma \bar{x}^2$  and  $\Sigma \bar{x}_i \bar{x}_j$  in terms of moments, he finds

$$(1.14) \quad \bar{\mu}_{2:z} = \frac{r(s-r)}{s-1} \bar{\mu}_{2:z}.$$

(It should be noted that Carver [12; p. 115] obtained the corresponding formula for infinite sampling by letting  $s \rightarrow \infty$ ).

The preceding development is entirely in accord with the first of the courses stated above. It is also the standard procedure and is the course followed by such writers as Isserles [2], Neyman [6], Church [7], Pepper [11] and Dwyer [20], in deriving finite sampling formulas. Also, it is the course followed by such authors as "Student" [1], Tchouproff [3], Church [5], Craig [9], Fisher [10], and Georgesque [13] for the case of sampling from an infinite population.



However, in (1.13), it is possible to employ the definition,

$$\frac{1}{s(s-1)} \sum \bar{x}_i \bar{x}_j = \bar{\mu}_{1,1}.$$

Then (1.14) becomes

$$(1.15) \quad \bar{\mu}_{2:s} = r\bar{\mu}_2 + r(r-1)\bar{\mu}_{1,1}.$$

Formula (1.15) may be interpreted as holding for either finite or infinite sampling, depending on the interpretation which is given to  $\bar{\mu}_{1,1}$ . It may be easily shown that, if the sampling is from a limited supply,  $\bar{\mu}_{1,1} = \frac{-1}{s-1} \bar{\mu}_2$  and (1.15) reduces to (1.14). If the sampling is from an infinite supply,  $\bar{\mu}_{1,1}$  becomes  $\bar{\mu}_1^2$  and therefore

$$\bar{\mu}_{2:s} = r\bar{\mu}_{2:s},$$

which is the formula [12; p. 115] that corresponds, in the infinite case, to (1.14).

Thus, either of the two courses is possible in the case of sampling from a single population. However, if one wishes to get general formulas which hold for both infinite and finite sampling, he should follow the second course. Similarly, in order to obtain generalized sampling formulas where the relations between the variates are unspecified and the populations are assumed to be different, the second course should be followed.

It appears that Tchouproff [3], [4] was the first to approach the sampling problem from such a general point of view. However, his methods of derivation are quite complicated and his results, in general, are difficult to apply to a given problem [5], [8].

Samples of  $n$  are formed from  ${}_nU_N$  by choosing one variate from each of the  $n$  populations. A typical sample is

$${}_1x_{i_1}, {}_2x_{i_2}, {}_3x_{i_3}, \dots, {}_rx_{i_r}, \dots, {}_nx_{i_n}.$$

We define [4; p. 472]

$$(1.16) \quad \frac{1}{k} \sum_{\substack{i_1, i_2, \dots, i_r=1 \\ r_j \neq r_k}}^n {}_1x_{i_1}^{t_1} {}_2x_{i_2}^{t_2} \dots {}_rx_{i_r}^{t_r} = E({}_1x_{i_1}^{t_1} {}_2x_{i_2}^{t_2} \dots {}_rx_{i_r}^{t_r}) \\ = {}_1r_2 \dots {}_rv \mu_{t_1 t_2 \dots t_v}, \\ \frac{v}{n^{(v)}} \sum_{r_j \neq r_k} {}_1r_2 \dots {}_rv \mu_{t_1 t_2 \dots t_v} = \frac{1}{n^{(v)}} S_v {}_1r_2 \dots {}_rv \mu_{t_1 t_2 \dots t_v} = \mu_{t_1 t_2 \dots t_v},$$

where  $k$  represents the number of possible terms of the given form;  $S_v$  means  $v$  times the sum for unequal values of  $r_1, r_2, \dots, r_v$  and  $n^{(v)} = n(n-1) \dots (n-v+1)$ .

**4. Moments and product moments of sample moments.** The  $t$ th moment of the  $j$ th sample is denoted by  ${}_jm_t$ . The  $s$ th moment of  ${}_jm_t$  for all  $j$  is denoted by  $'\mu_{s:m_t}$ , where the prime indicates that the moments of the universe are measured about a fixed point. It follows that

$$(1.17) \quad {}_j m_i = \frac{1}{n} \sum_{r=1}^n r x_{i,r}^i \quad \text{and} \quad {}'_\mu{}_{2:m_i} = E[{}_j m_i]^2.$$

Also, the general product moment, in which the variates of both the sample and the universe are measured about a fixed point, is defined by

$$(1.18) \quad {}'_\mu{}_{s_1 s_2 \dots s_r : m_{i_1} m_{i_2} \dots m_{i_r}} = E[{}_j m_{i_1}^{s_1} {}_j m_{i_2}^{s_2} \dots {}_j m_{i_r}^{s_r}].$$

As an illustration of the methods used to derive the formulas of this section, consider a special case of (1.18) when  $s_1 = 2$  and  $s_i = 0$ , ( $i = 2, 3, \dots, v$ ). Then

$$\begin{aligned} {}'_\mu{}_{2:m_i} &= \frac{1}{n^2} E \left[ \sum_{r=1}^n r x_{i,r}^i \right]^2 \\ &= \frac{1}{n^2} E \left[ \sum_{r=1}^n r x_{i,r}^{2i} + S_{2,r_1} x_{i,r_1}^i r_2 x_{i,r_2}^i \right] \\ &= \frac{1}{n^2} \left[ \sum_{r=1}^n r \mu_{2i} + S_{2,r_1 r_2} \mu_{i,i} \right]. \end{aligned}$$

Therefore, by (1.1), (1.2) and (1.16), we get

$$(1.19) \quad {}'_\mu{}_{2:m_i} = \frac{1}{n^2} [n \mu_{2i} + n^{(2)} \mu_{i,i}].$$

Using the formulas [20; p. 34] relating products of power sums and power products to expand expressions of the type  $E({}_j m_{i_1}^{s_1} {}_j m_{i_2}^{s_2} \dots {}_j m_{i_r}^{s_r})$ , we give, in the tables below, formulas for moments and product moments of sample moments through weight six. The number in a cell and the coefficient, in the same column, at the top of the table should be taken as the coefficient of the moment which is found in the same vertical division. The coefficients in the vertical division are coefficients of the entire right members of the formulas for the respective moments.

Terms of the form  $\mu_{i_1 i_2 \dots i_r}$ , if  $i_1 = i_2 = \dots = i_r = i$ , are sometimes written  $\mu_{ir, i_r+1 \dots i_r}$ .

The numbers in the cells of the tables are identical with the numbers in the cells of the tables given by Dwyer [19; p. 30] for the expected value of partition products.

**5. Moments of central moments of samples of  $n$ .** The  $t$ th central moment of the  $j$ th sample is denoted by  ${}_j \bar{m}_t$ . Then,

$$(1.20) \quad {}_j \bar{m}_t = \frac{1}{n} \sum_{r=1}^n (x_{i,r} - {}_j m_i)^t$$

and

$$(1.21) \quad {}'_\mu{}_{2:\bar{m}_i} = E \left[ \frac{1}{n} \sum_{r=1}^n (x_{i,r} - {}_j m_i)^2 \right]^2.$$

TABLE I

(1)				(2)				(3)			
Coef.				Coef.				Coef.	$n$	$n^{(2)}$	$n^{(3)}$
$\mu_1$				$\mu_2 \quad \mu_1^2$					$\mu_2$	$\mu_{2,1}$	$\mu_1^3$
$'\mu_{1:m_1}$				$'\mu_{1:m_2} \quad n^{-1} \quad 1$				$'\mu_{1:m_3}$	$n^{-1}$	1	
				$'\mu_{2:m_1} \quad n^{-2} \quad 1$				$'\mu_{11:m_1m_2}$	$n^{-2}$	1	1
								$'\mu_{3:m_1}$	$n^{-3}$	1	3
											1

(5)								
Coef.	$n$	$n^{(2)}$	$n^{(3)}$	$n^{(3)}$	$n^{(4)}$	$n^{(5)}$	$n^{(6)}$	
	$\mu_6$	$\mu_{4,1}$	$\mu_{3,2}$	$\mu_{3,1^2}$	$\mu_{2^2,1}$	$\mu_{2,1^2}$	$\mu_1^3$	
$'\mu_{1:m_3}$	$n^{-1}$	1						
$'\mu_{11:m_1m_4}$	$n^{-2}$	1	1					
$'\mu_{11:m_2m_3}$	$n^{-2}$	1		1				
$'\mu_{21:m_1m_2}$	$n^{-3}$	1	2	1	1			
$'\mu_{12:m_1m_2}$	$n^{-3}$	1	1	2		1		
$'\mu_{31:m_1m_2}$	$n^{-4}$	1	3	4	3	3	1	
$'\mu_{6:m_1}$	$n^{-5}$	1	5	10	10	15	10	1

(4)						
Coef.	$n$	$n^{(2)}$	$n^{(2)}$	$n^{(3)}$	$n^{(4)}$	
	$\mu_4$	$\mu_{3,1}$	$\mu_{2,2}$	$\mu_{2,1^2}$	$\mu_1^4$	
$'\mu_{1:m_4}$	$n^{-1}$	1				
$'\mu_{11:m_1m_3}$	$n^{-2}$	1	1			
$'\mu_{2:m_2}$	$n^{-2}$	1		1		
$'\mu_{21:m_1m_2}$	$n^{-3}$	1	2	1	1	
$'\mu_{4:m_1}$	$n^{-4}$	1	4	3	6	1

(6)												
Coef.	$n$	$n^{(2)}$	$n^{(2)}$	$n^{(2)}$	$n^{(3)}$	$n^{(3)}$	$n^{(3)}$	$n^{(4)}$	$n^{(4)}$	$n^{(5)}$	$n^{(6)}$	
	$\mu_6$	$\mu_{5,1}$	$\mu_{4,2}$	$\mu_{3,3}$	$\mu_{4,1^2}$	$\mu_{3,2,1}$	$\mu_2^3$	$\mu_{3,1^3}$	$\mu_{2^2,1^2}$	$\mu_{2,1^4}$	$\mu_1^6$	
$'\mu_{1:m_6}$	$n^{-1}$	1										
$'\mu_{11:m_1m_5}$	$n^{-2}$	1	1									
$'\mu_{11:m_2m_4}$	$n^{-2}$	1		1								
$'\mu_{2:m_3}$	$n^{-2}$	1			1							
$'\mu_{21:m_1m_4}$	$n^{-3}$	1	2	1		1						
$'\mu_{111:m_1m_2m_3}$	$n^{-3}$	1	1	1	1		1					
$'\mu_{3:m_2}$	$n^{-3}$	1		3			1					
$'\mu_{31:m_1m_3}$	$n^{-4}$	1	3	3	1	3	3	1				
$'\mu_{32:m_1m_2}$	$n^{-4}$	1	2	3	2	1	4	1	1			
$'\mu_{41:m_1m_3}$	$n^{-5}$	1	4	7	4	6	16	3	4	6	1	
$'\mu_{6:m_1}$	$n^{-6}$	1	6	15	10	15	60	15	20	45	15	1

After writing  $(x_{i_r} - i_r m_1)^t$  as the sum of the general term of a binomial series and then expanding the resulting right member of (1.21) as a product of power sums [20; p. 19], we get

$$(1.22) \quad \begin{aligned} \mu_{t:\bar{m}_t} = \sum \frac{s!}{r_1! r_2! \dots r_v! \pi_1! \pi_2! \dots} \sum_{\substack{i_1, i_2, \dots, i_v=0 \\ i_1 r_1 + i_2 r_2 + \dots + i_v r_v = t}} (-1)^\rho \binom{t}{i_1}^{r_1} \binom{t}{i_2}^{r_2} \\ \dots \binom{t}{i_v}^{r_v} \mu_{r_1 r_2 \dots r_v \rho: m_1 - i_1 m_1 - i_2 m_1 - \dots - i_v m_1} \end{aligned}$$

where  $\sum_{j=1}^v r_j = s$ ,  $\sum_{j=1}^v i_j r_j = \rho$  and  $\pi_1, \pi_2, \dots$  are the numbers of the repeated parts of  $s$ .

The mean of the  $t$ th central moment takes the following simple form,

$$(1.23) \quad \mu_{1:\bar{m}_t} = \sum_{i=0}^t (-1)^i \binom{t}{i} \mu_{1i: m_1 - i m_1},$$

where the moments in the right member of (1.23) through weight six are given in the tables of section four. Also,

$$(1.24) \quad \mu_{2:\bar{m}_2} = \mu_{2: m_2} - 2' \mu_{21: m_1 m_2} + \mu_{4: m_1}.$$

$$(1.25) \quad \mu_{3:\bar{m}_2} = \mu_{3: m_2} - 3' \mu_{22: m_1 m_2} + 3' \mu_{41: m_1 m_2} - \mu_{6: m_1}.$$

$$(1.26) \quad \begin{aligned} \mu_{2:\bar{m}_3} = \mu_{2: m_3} + 9' \mu_{22: m_1 m_2} + 4' \mu_{6: m_1} - 6' \mu_{111: m_1 m_2 m_3} \\ + 4' \mu_{31: m_1 m_2} - 12' \mu_{41: m_1 m_2}. \end{aligned}$$

After substituting from the tables of section four, (1.23) through (1.26) become

$$(1.27) \quad \mu_{1:\bar{m}_2} = \frac{n^{(2)}}{n^2} [\mu_2 - \mu_{1,1}].$$

$$(1.28) \quad \mu_{1:\bar{m}_3} = \frac{n^{(3)}}{n^3} [\mu_3 - 3\mu_{2,1} + 2\mu_{1,2}].$$

$$(1.29) \quad \begin{aligned} \mu_{1:\bar{m}_4} = \frac{1}{n^4} [n^{(2)}(n^2 - 3n + 3)(\mu_4 - 4\mu_{3,1}) + 3n^{(2)}(2n - 3)\mu_{2,2} \\ + 3n^{(4)}(2\mu_{2,1^2} - \mu_{1,4})]. \end{aligned}$$

$$(1.30) \quad \begin{aligned} \mu_{1:\bar{m}_5} = \frac{1}{n^5} [n^{(3)}(n^2 - 2n + 2)(\mu_5 - 5\mu_{4,1}) + 10n^{(3)}(n - 2)\mu_{3,2} \\ + 10n^{(3)}(n + 1)(n - 4)\mu_{3,1^2} - 30n^{(3)}(n - 2)\mu_{2^2,1} \\ - 10n^{(4)}(3n - 4)\mu_{2,1^3} + 4n^{(5)}\mu_{1,5}]. \end{aligned}$$

$$\begin{aligned}
 {}'\mu_{1;\bar{m}_6} &= \frac{1}{n^6} [n^2(n^4 - 5n^3 + 10n^2 - 10n + 5)(\mu_6 - \mu_{5,1}) \\
 &\quad + 15n^{(2)}(n^3 - 4n^2 + 7n - 5)\mu_{4,2} - 10n^{(3)}(2n^2 - 6n + 5)\mu_{3,1} \\
 &\quad + 15n^{(3)}(n^3 - 4n^2 + 6n - 5)\mu_{4,1,1} - 60n^{(3)}(n^2 - 4n + 5)\mu_{3,2,1} \\
 &\quad + 15n^{(3)}(3n - 5)\mu_{2,1} - 20n^{(4)}(n^2 - 3n + 5)\mu_{3,1,1} \\
 &\quad + 45n^{(4)}(2n - 5)\mu_{2,1,1} + 15n^{(5)}(n - 5)\mu_{2,1,1} - 5n^{(6)}\mu_{1,6}].
 \end{aligned}
 \tag{1.31}$$

$${}'\mu_{2;\bar{m}_2} = \frac{1}{n^4} [n^{(2)}(n - 1)(\mu_4 - 4\mu_{3,1}) + n^{(3)}(n + 1)\mu_{2,2} - n^{(4)}(2\mu_{2,1,1} - \mu_{1,4})].
 \tag{1.32}$$

$$\begin{aligned}
 {}'\mu_{3;\bar{m}_2} &= \frac{1}{n^6} [n^{(2)}(n - 1)^2(\mu_6 - 6\mu_{5,1}) + 3n^{(2)}(n - 1)(n^2 - 2n + 5)\mu_{4,2} \\
 &\quad - 2n^{(2)}(3n^2 - 6n + 5)\mu_{3,3} + n^{(3)}(n^3 - 3n^2 + 9n - 15)\mu_{2,1} \\
 &\quad - 3n^{(3)}(n - 1)(n - 5)\mu_{4,1,1} - 12n^{(3)}(n^2 - 4n + 5)\mu_{3,2,1} \\
 &\quad + 4n^{(4)}(3n - 5)\mu_{3,1,1} - 3n^{(4)}(n^2 - 6n + 15)\mu_{2,1,1} \\
 &\quad + n^{(6)}(3\mu_{2,1,1} - \mu_{1,6})].
 \end{aligned}
 \tag{1.33}$$

$$\begin{aligned}
 {}'\mu_{2;\bar{m}_3} &= \frac{1}{n^6} [n^{(2)}(n - 1)^2(n - 2)(\mu_6 - 6\mu_{5,1}) - 3n^{(2)}(n - 2)^2(2n - 5)\mu_{4,2} \\
 &\quad + n^{(2)}(n - 2)^2(n^2 - 2n + 10)\mu_{3,3} \\
 &\quad - 6n^{(3)}(n - 2)(n^2 - 6n + 20)\mu_{3,2,1} + 3n^{(3)}(n - 2)(7n - 10)\mu_{4,1,1} \\
 &\quad + 3n^{(3)}(3n^2 - 12n + 20)\mu_{2,1} + 4n^{(4)}(n - 2)(n - 10)\mu_{3,1,1} \\
 &\quad + 9n^{(4)}(n^2 - 8n + 20)\mu_{2,1,1} - 4n^{(6)}(3\mu_{2,1,1} - \mu_{1,6})].
 \end{aligned}
 \tag{1.34}$$

**6. The variance of the variance of samples of  $n$ .** The variance of the variance of samples of  $n$ , when the moments of the universe are measured about a fixed point, is defined as

$${}'\bar{\mu}_{2;\bar{m}_2} = {}'\mu_{2;\bar{m}_2} - [{}'\mu_{1;\bar{m}_2}]^2.
 \tag{1.35}$$

Therefore, from (1.27) and (1.32),

$$\begin{aligned}
 {}'\mu_{2;\bar{m}_2} &= \frac{1}{n^4} [n^{(2)}(n - 1)(\mu_4 - 4\mu_{3,1}) + n^{(3)}(n - 1)\mu_{2,2} - n^{(4)}(2\mu_{2,1,1} - \mu_{1,4})] \\
 &\quad - \left( \frac{n - 1}{n} \right)^2 (\mu_2 - \mu_{1,1})^2.
 \end{aligned}
 \tag{1.36}$$

Tchouproff [4; p. 492] gave a formula (8) for the variance of the sample variance but his result is unwieldy due to the fact that moments of the universe are measured about the mean.

**7. Conventional infinite sampling formulas derived from generalized sampling formulas.** The term "infinite sampling" is to be interpreted as meaning: *sampling from an unlimited supply or sampling from a limited supply with repetitions permitted.* In each of these situations the variates are independent [5; p. 79].

First, it is assured that the  $n$  populations are identical, that is,  ${}_1X = {}_2X = \dots = {}_nX$ . This assumption results in the fact that, for a fixed  $t$ ,  ${}_1\mu_t = {}_2\mu_t = \dots = {}_n\mu_t$  and  ${}_1\bar{\mu}_t = {}_2\bar{\mu}_t = \dots = {}_n\bar{\mu}_t$ . Therefore, under the assumption of identical populations, every moment may be interpreted as either the moment of  $n$  identical populations or as the moment of a single population. The only other assumption is that the sampling is "infinite".

From the condition of independence [3; p. 141], we have

$$E(r_1 x_{i_1}^{t_1} r_2 x_{i_2}^{t_2} \dots r_v x_{i_v}^{t_v}) = (E r_1 x_{i_1}^{t_1}) (E r_2 x_{i_2}^{t_2}) \dots (E r_v x_{i_v}^{t_v}).$$

Therefore,

$$r_1 r_2 \dots r_v \mu_{t_1 t_2 \dots t_v} = r_1 \mu_{t_1} r_2 \mu_{t_2} \dots r_v \mu_{t_v}.$$

Combining the condition of independence with that of identical populations, we have

$$(1.37) \quad \frac{1}{n^{(v)}} S_v r_1 r_2 \dots r_v \mu_{t_1 t_2 \dots t_v} = \frac{1}{n^{(v)}} S_v r_1 \mu_{t_1} r_2 \mu_{t_2} \dots r_v \mu_{t_v} = \mu_{t_1} \mu_{t_2} \dots \mu_{t_v}.$$

By (1.16) and (1.37), we may write

$$(1.38) \quad \mu_{t_1 t_2 \dots t_v} = \mu_{t_1} \mu_{t_2} \dots \mu_{t_v}.$$

Since the only terms of the generalized sampling formulas which are affected by the assumption of "infinite sampling" are those of the form  $\mu_{t_1 t_2 \dots t_v}$ , the problem of obtaining conventional infinite sampling formulas from generalized sampling formulas is, in practice, a mechanical one. Simply write terms of the form  $\mu_{t_1 t_2 \dots t_v}$  which appear in a generalized sampling formula, as  $\mu_{t_1} \mu_{t_2} \dots \mu_{t_v}$ , and one automatically obtains the corresponding infinite sampling formula.

As an illustration of the method, consider the generalized sampling formula (1.36) for the variance of the sample variance. When (1.38) is utilized to change it into the corresponding infinite sampling formula, (1.36) becomes

$$(1.39) \quad \bar{\mu}_{2;\bar{m}_2} = \frac{n^{(2)}}{n^4} [(n-1)(\mu_4 - 4\mu_3\mu_1) - (n-3)\mu_2^2 + 2(2n-3)(2\mu_2\mu_1^2 - \mu_1^4)],$$

which is the usual formula [20; p. 75] for the variance of the sample variance when the moments of the universe are measured about a fixed point. If it is assumed that the moments of  ${}_nU_N$  are measured about the mean, formula (1.39) becomes

$$(1.40) \quad \bar{\mu}_{2;\bar{m}_2} = \frac{n^{(2)}}{n^4} [(n-1)\bar{\mu}_4 - (n-3)\bar{\mu}_2^2],$$

which was published by "Student" [1; p. 3] in 1908.

**8. Conventional finite sampling formulas derived from generalized sampling formulas.** The term "finite sampling" is to be interpreted as meaning: *sampling from a limited supply when repetitions are not permitted.*

In order to reduce generalized sampling formulas to the corresponding formulas for finite sampling, the assumptions are made that the  $n$  populations are identical and that  $N$  and  $n$  are finite,  $N > n$ . The selection of variates which enter each sample is restricted in the following manner. If a variate having a given post-subscript is chosen, then no other variate having the same post-subscript may be chosen for the same sample.

Now it is evident that terms of the form  $\mu_{t_1 t_2 \dots t_v}$  must be redefined on the basis of the preceding assumptions. From the expansions [20; p. 32] of power product sums in terms of products of power sums, we get the formulas for  $\mu_{t_1 t_2 \dots t_v}$  which are given in the following tables.

The formulas in the tables of this section are called *transformation formulas for finite sampling* or more briefly *transformation formulas*.

The transformation of generalized sampling formulas into corresponding finite sampling formulas is illustrated by the substitution of  $\frac{N^2 \mu_1^2 - N \mu_2}{N^{(2)}}$  for  $\mu_{1,1}$  in (1.27). We get

$$(1.41) \quad \mu_{1;\bar{m}_2} = \frac{N(n-1)}{n(N-1)} [\mu_2 - \mu_1^2],$$

which is the well-known finite sampling formula for the mean of the variance of samples of  $n$ .

From this and the preceding section it is evident that the generalized sampling formulas may be considered as formulas for either infinite or finite sampling depending upon the interpretation given to terms of the form  $\mu_{t_1 t_2 \dots t_v}$ .

**9. Transformation of infinite sampling formulas into corresponding finite sampling formulas.** It is a well-known fact that infinite sampling formulas may be obtained from those for finite sampling by letting the size of the parent population become infinite. But, prior to this paper, apparently no one has presented a method of obtaining finite sampling formulas from infinite sampling formulas. However, by making use of the relations between finite, infinite, and generalized sampling, we shall demonstrate that it is possible to transform any infinite sampling formula into the corresponding finite sampling formula.

Since the infinite sampling formulas are obtained from the generalized sampling formulas by replacing

$$\mu_{t_1 t_2 \dots t_v} \text{ by } \mu_{t_1} \mu_{t_2} \dots \mu_{t_v}$$

it follows that generalized sampling formulas may be obtained from the infinite

TABLE II

(1)			(2)				(3)				
	Coef.	$N$		Coef.	$N$	$N^2$		Coef.	$N$	$N^2$	$N^3$
		$\mu_1$			$\mu_2$	$\mu_1^2$			$\mu_3$	$\mu_2\mu_1$	$\mu_1^3$
$\mu_1$	$N^{(-1)}$	1	$\mu_2$	$N^{(-1)}$	1		$\mu_3$	$N^{(-1)}$	1		
			$\mu_{1,1}$	$N^{(-2)}$	-1	1	$\mu_{2,1}$	$N^{(-2)}$	-1	1	
							$\mu_{1^2}$	$N^{(-2)}$	2	-3	1

(5)								
	Coef.	$N$	$N^2$	$N^3$	$N^4$	$N^5$	$N^6$	$N^7$
		$\mu_5$	$\mu_4\mu_1$	$\mu_3\mu_2$	$\mu_2\mu_1^2$	$\mu_1^2\mu_1$	$\mu_2\mu_1^2$	$\mu_1^3$
$\mu_5$	$N^{(-1)}$	1						
$\mu_{4,1}$	$N^{(-2)}$	-1	1					
$\mu_{3,2}$	$N^{(-2)}$	-1		1				
$\mu_{2,1^2}$	$N^{(-2)}$	2	-2	-1	1			
$\mu_{2^2,1}$	$N^{(-2)}$	2	-1	-2		1		
$\mu_{2,1^3}$	$N^{(-4)}$	-6	6	5	-3	-3	1	
$\mu_{1^5}$	$N^{(-5)}$	24	-30	-20	20	15	-10	1

(4)							
	Coef.	$N$	$N^2$	$N^3$	$N^4$	$N^5$	$N^6$
		$\mu_4$	$\mu_{3,1}$	$\mu_2^2$	$\mu_2\mu_1^2$	$\mu_1^3$	
$\mu_4$	$N^{(-1)}$	1					
$\mu_{3,1}$	$N^{(-2)}$	-1	1				
$\mu_{2,2}$	$N^{(-2)}$	-1		1			
$\mu_{2,1^2}$	$N^{(-2)}$	2	-2	-1	1		
$\mu_{1^4}$	$N^{(-4)}$	-6	8	3	-6	1	

(6)												
	Coef.	$N$	$N^2$	$N^3$	$N^4$	$N^5$	$N^6$	$N^7$	$N^8$	$N^9$	$N^{10}$	$N^{11}$
		$\mu_6$	$\mu_5\mu_1$	$\mu_4\mu_2$	$\mu_3^2$	$\mu_4\mu_1^2$	$\mu_3\mu_2\mu_1$	$\mu_2^3$	$\mu_3\mu_1^2$	$\mu_1^2\mu_1^2$	$\mu_2\mu_1^3$	$\mu_1^4$
$\mu_6$	$N^{(-1)}$	1										
$\mu_{5,1}$	$N^{(-2)}$	-1	1									
$\mu_{4,2}$	$N^{(-2)}$	-1		1								
$\mu_{3,3}$	$N^{(-2)}$	-1			1							
$\mu_{4,1^2}$	$N^{(-2)}$	2	-2	-1		1						
$\mu_{3,2,1}$	$N^{(-2)}$	2	-1	-1	-1		1					
$\mu_{2^3}$	$N^{(-2)}$	2		-3				1				
$\mu_{3,1^3}$	$N^{(-4)}$	-6	6	3	2	-3	-3		1			
$\mu_{2^2,1^2}$	$N^{(-4)}$	-6	4	5	2	-1	-4	-1		1		
$\mu_{2,1^4}$	$N^{(-6)}$	24	-24	-18	-8	12	20	3	-4	-6	1	
$\mu_{1^6}$	$N^{(-6)}$	-120	144	90	40	-90	-120	-15	40	45	-15	1



formulas by replacing

$$(1.42) \quad \mu_{t_1} \mu_{t_2} \cdots \mu_{t_v} \text{ by } \mu_{t_1 t_2 \cdots t_v}.$$

However, it must be emphasized that the application of (1.42) demands formulae which are expressed in terms of moments of sample moments rather than central moments of sample moments (although the sample moments may be measured about a fixed point or about the mean) and the moments of the universe must be measured about a fixed point. The reason for these restrictions is to insure that each term is accounted for individually.

After replacements (1.42) are made in the formula for sampling from an infinite population, the resulting formula is the corresponding generalized one. The step to the corresponding finite sampling formulas is simply the one outlined in section eight, namely, the use of the transformation formulas.

We shall consider, as the first illustration, the infinite sampling formula for the mean of the sample variance when the moments of the parent population are measured about the mean. The formula is

$$(1.43) \quad \mu_{1:\bar{m}_2} = \frac{n-1}{n} \bar{\mu}_2.$$

When (1.43) is expressed in terms of moments of the parent population about a fixed point, we have

$$(1.44) \quad \mu_{1:\bar{m}_2} = \frac{n-1}{n} [\mu_2 - \mu_1^2].$$

Following (1.42),  $\mu_1^2$  is replaced by  $\mu_{1,1}$  and (1.44) becomes (1.27). The use of the transformation formula for  $\mu_{1,1}$  gives (1.41) which, when the moments of the parent population are measured about the mean, becomes

$$(1.45) \quad \mu_{1:\bar{m}_2} = \frac{N(n-1)}{n(N-1)} \bar{\mu}_2.$$

Infinite sampling formulas expressed in terms of moment-function, may be similarly transformed into the corresponding finite sampling formulas. For example, Craig [9; p. 57] gives the second Thiele seminvariant of the variance of samples as

$$(1.46) \quad \lambda_{2:\bar{m}_2} = \frac{(n-1)^2}{n^3} \lambda_4 + 2 \frac{(n-1)}{n^2} \lambda_2^2.$$

First, we express (1.46) in terms of moments about a fixed point by use of the formulas relating Thiele seminvariants and moments [9; p. 12]. We also recall that the resulting formula should be expressed in terms of moments of sample moments rather than in terms of central moments of sample moments. We obtain

$$(1.47) \quad \begin{aligned} \mu_{2:\bar{m}_2} = \frac{(n-1)}{n^3} [(n-1)\mu_4 - 4(n-1)\mu_3\mu_1 + (n^2 - 2n + 3)\mu_2^2 \\ - 2(n-2)(n-3)\mu_2\mu_1^2 + (n-2)(n-3)\mu_1^4]. \end{aligned}$$

The next step is to transform (1.47) into the corresponding generalized sampling formula by use of (1.42). We obtain (1.32). Since we desire to obtain the finite sampling formula which exactly corresponds to (1.46), it is necessary to transform (1.32) from the second moment of  $\bar{m}_2$  to the variance of  $\bar{m}_2$  and we get (1.36). Next the transformation formulas are applied to (1.36). When the moments of the parent population are measured about the mean and are replaced by Thiele seminvariants, (1.36) becomes

$$(1.48) \quad \lambda_{2:\bar{m}_2} = \frac{N(N-n)(n-1)}{n^2(N-1)^2(N-2)(N-3)} [(N-1)(Nn-N-n-1)\lambda_4 \\ + 2(N^2n-3Nn-3N+3n+3)\lambda_2^2].$$

Formula (1.48) gives the second Thiele seminvariant of the variance of samples of  $n$  drawn from a finite parent population of  $N$ . When  $N \rightarrow \infty$ , in (1.48), we obtain immediately (1.46).

It is generally true that infinite sampling formulas are more easily derived than are the corresponding finite sampling formulas. The methods of this section make it possible to derive the desired sampling formulas for the infinite parent population and then transform these infinite sampling formulas into the corresponding finite sampling formulas.

## II. MOMENT FUNCTION ADJUSTMENTS FOR GROUPED DATA

A given distribution of discrete variates may be grouped in " $k$  groupings of  $k$ ". We desire to find the correction which eliminates the error made in replacing a given moment of the original distribution by the average of the corresponding moments of the  $k$  grouped-distributions.

Formulas for the adjustments for moments of a grouped-distribution of discrete variates were first given (without proof) in the Editorial of Vol. I, No. 1 of the *Annals of Mathematical Statistics*. Later, more satisfactory derivations of adjustment formulas were given by Abernethy [24] Craig [25] and Carver [26]. However, it was observed by Carver [26; p. 162] that the developments of Abernethy and Craig are adjustments about a fixed point and that they fail to hold for the case of expectations of central moments if we accept the definition

$$\mu_{1:\bar{\mu}_t} = \frac{1}{k} \sum_{r=1}^k r\bar{\mu}_t, \quad (t = 2, 3, \dots).$$

Here  $r\bar{\mu}_t$  represents the  $t$ th central moment of the  $r$ th grouped-distribution. The formula for the true value of  $\mu_{1:\bar{\mu}_2}$  was supplied by Carver [26; p. 162] but he did not indicate a general method which might be used for the derivation of  $\mu_{1:\bar{\mu}_t}$ ; ( $t > 2$ ).

A distribution of discrete variates grouped in " $k$  groupings of  $k$ " is a special case of a universe of  $n$  finite populations and hence the methods and formulas for the expectations of population moments are applicable to our present problem.

It is found that the adjustment formulas for moment-functions of grouped data involve central moments of a rectangular distribution. It will be convenient for our present purposes to give a brief treatment of the moment-functions of a rectangular distribution.

**1. Moment-functions of a rectangular distribution.** Consider the rectangular distribution of discrete variates,

$$(2.1) \quad h, 2h, 3h, \dots, kh.$$

It is readily shown that the moment generating function of (2.1),

$$(2.2) \quad G_x(\theta) = \mu_0 + \mu_1\theta + \mu_2 \frac{\theta^2}{2!} + \dots + \mu_n \frac{\theta^n}{n!} + \dots$$

may be written

$$(2.3) \quad G_x(\theta) = \frac{e^{i(k+1)h\theta} \sinh \frac{1}{2}kh\theta}{k \sinh \frac{1}{2}h\theta}.$$

Setting the expansion of the right member of (2.3) equal to the right member of (2.2) and equating coefficients of like powers of  $\theta$ , we obtain the following recursion formula for the moments of (1.1)

$$(2.4) \quad \frac{(n+1)^{(1)}}{1!} \mu_{n:R} - \frac{(n+1)^{(2)}}{2!} h \mu_{n-1:R} + \dots + (-1)^{r-1} \frac{(n+1)^{(r)}}{r!} h^{r-1} \mu_{n-r+1:R} + \dots = k^n h^n,$$

where  $\mu_{n:R}$  represents the  $n$ th moment of a rectangular distribution. Formulas for  $\mu_{n:R}$ , ( $n = 0, 1, \dots, 10$ ) are given below. See Sasuly [27; p. 27].

$$\mu_{0:R} = 1.$$

$$\mu_{1:R} = \frac{1}{2}(k+1)h.$$

$$\mu_{2:R} = \frac{1}{6}(k+1)(2k+1)h^2 = \frac{1}{3}(2k+1)h\mu_{1:R}.$$

$$\mu_{3:R} = \frac{1}{4}(k+1)^2 kh^3 = kh\mu_{1:R}^2.$$

$$\mu_{4:R} = \frac{1}{5}(3k^2 + 3k - 1)h^2 \mu_{2:R}.$$

$$(2.5) \quad \mu_{5:R} = \frac{1}{3}(2k^2 + 2k + 1)h^2 \mu_{3:R}.$$

$$\mu_{6:R} = \frac{1}{7}(3k^4 + 6k^3 - 3k + 1)h^4 \mu_{2:R}.$$

$$\mu_{7:R} = \frac{1}{6}(3k^4 + 6k^3 - k^2 - 4k + 2)h^4 \mu_{3:R}.$$

$$\mu_{8:R} = \frac{1}{15}(5k^6 + 15k^5 + 5k^4 - 15k^3 - k^2 + 9k - 3)h^6 \mu_{2:R}.$$

$$\mu_{9:R} = \frac{1}{5}(2k^6 + 6k^5 + k^4 - 8k^3 + k^2 + 6k - 3)h^6 \mu_{3:R}.$$

$$\mu_{10:R} = \frac{1}{11}(3k^8 + 12k^7 + 8k^6 - 18k^5 - 10k^4 + 24k^3 + 2k^2 - 15k + 5)h^8 \mu_{2:R}.$$

The deviations about the mean of (2.1) are

$$(2.6) \quad -\frac{1}{2}(k-1)h, \quad -\frac{1}{2}(k-3)h, \dots, \frac{1}{2}(k-3)h, \quad \frac{1}{2}(k-1)h.$$

Therefore,

$$(2.7) \quad \bar{\mu}_{2n+1;R} = 0.$$

If we denote (2.6) by  $\bar{x}$ , we have

$$(2.8) \quad G_2(\theta) = \frac{\sinh \frac{1}{2}(kh\theta)}{k \sinh \frac{1}{2}(h\theta)}.$$

The recursion formula for central moments of (2.1) is

$$(2.9) \quad \frac{(2n+1)^{(1)}}{1!} \bar{\mu}_{2n;R} + \frac{h^2 (2n+1)^{(3)}}{2^2 \cdot 3!} \bar{\mu}_{2n-2;R} + \dots + \frac{h^r (2n+1)^{(r+1)}}{2^r (r+1)!} \bar{\mu}_{2n-r;R} + \dots = \frac{k^{2n} h^{2n}}{2^{2n}}.$$

Formulas for  $\bar{\mu}_{2n;R}$ , ( $n = 0, 1, \dots, 5$ ) are given below. See [27; p. 27].

$$(2.10) \quad \begin{aligned} \bar{\mu}_{0;R} &= 1, \\ \bar{\mu}_{2;R} &= \frac{1}{12}(k^2 - 1)h^2, \\ \bar{\mu}_{4;R} &= \frac{1}{240}(3k^2 - 7)h^2 \bar{\mu}_{2;R}, \\ \bar{\mu}_{6;R} &= \frac{1}{1152}(3k^4 - 18k^2 + 31)h^4 \bar{\mu}_{2;R}, \\ \bar{\mu}_{8;R} &= \frac{1}{98304}(5k^6 - 55k^4 + 239k^2 - 381)h^6 \bar{\mu}_{2;R}, \\ \bar{\mu}_{10;R} &= \frac{1}{2814720}(3k^8 - 52k^6 + 410k^4 - 1636k^2 + 2555)h^8 \bar{\mu}_{2;R}. \end{aligned}$$

From the relation which connects Thiele seminvariants and the moment generating function, we get, see [25; p. 57],

$$(2.11) \quad \begin{aligned} \lambda_{0;R} &= 0, \quad \lambda_{1;R} = \frac{(k+1)h}{2}, \quad \lambda_{2n+1;R} = 0, \\ \lambda_{2n;R} &= (-1)^{n+1} \frac{B_n h^{2n} (k^{2n} - 1)}{2n}, \quad n = 1, 2, 3, \dots \end{aligned}$$

where  $\lambda_{n;R}$  represents the  $n$ th Thiele seminvariant of a rectangular distribution of discrete variates and  $B_n$ , ( $n = 1, 2, \dots$ ), the Bernoulli numbers:  $\frac{1}{6}, \frac{1}{30}, \dots$ .

In each of the cases considered in this section, corresponding formulas may be found for a rectangular distribution of continuous variates by setting  $h = m/k$  (which makes the range  $m$  with  $k$  subdivisions) and then letting  $k \rightarrow \infty$ .

**2. Adjustments for moments.** As our basic distribution we consider the set of discrete variates,  $x_i$ , ( $i = 1, 2, \dots, N$ ), where some of the  $x_i$ 's may not be distinct. We assume that the given distribution is grouped in " $k$  groupings of  $k$ ".

When  $x_i$  is placed in the  $r$ th position of a class, the limits of the class are  $x_i - (r - 1)h$  and  $x_i + (k - r)h$  and the class mark is  $x_i + \left[ \frac{k - (2r - 1)}{2} \right] h$ . Thus, when the class mark is used as the value of  $x_i$ , the quantity  $\left[ \frac{k - (2r - 1)}{2} \right] h$  is added to the true value of  $x_i$ . Therefore, when the expected value of a particular moment for " $k$  groupings of  $k$ " is found, each variate has made a definite contribution as it was placed in each of the  $k$  positions of a class.

For convenience, we define

$$(2.12) \quad e_r = \left[ \frac{k - (2r - 1)}{2} \right] h, \quad (r = 1, 2, \dots, k).$$

As was previously indicated, the expected value of a given moment involves the contribution of each variate as it occupies the  $k$  class positions. A convenient method of finding these contributions is by means of a universe  ${}_kU_N$  which is composed of the populations  ${}_rX$ , ( $r = 1, 2, \dots, k$ ). The  $r$ th population consists of the values of the variates when they occupy the  $r$ th position of the class. Hence  ${}_rX$  consists of  ${}_rx_i = x_i + e_r$ , ( $i = 1, 2, \dots, N$ ).

The notation for moments is the same as that of Part I. Since  ${}_kU_N$  is of the same form as the universe studied in Part I, we use the definitions (1.1) of that part.

The expected value of the  $t$ th moment is

$$\begin{aligned} \mu_{1:\mu_t} &= \frac{1}{k} \sum_{r=1}^k E(x_i + e_r)^t \\ &= \sum_{s=0}^t \binom{t}{s} \left[ \frac{1}{k} \sum_{r=1}^k e_r^s \right] \mu_{t-s}. \end{aligned}$$

Many devices have been used by previous writers [24; p. 269], [25; p. 57], [26; p. 157], to evaluate terms of the form  $\frac{1}{k} \sum_{r=1}^k e_r^s$ . However, it should be noticed that the quantities  $e_r$ , ( $i = 1, 2, \dots, k$ ), are respectively identical with the deviations (2.6) about the mean of a rectangular distribution of discrete variates. It follows that

$$\bar{\mu}_{s;R} = \frac{1}{k} \sum_{r=1}^k e_r^s.$$

And since  $\bar{\mu}_{2s+1;R} = 0$ , we have

$$(2.13) \quad \mu_{1:\mu_t} = \sum_{s=0}^{\lfloor t/2 \rfloor} \binom{t}{2s} \mu_{t-2s} \bar{\mu}_{2s;R}.$$

Formulas for  $\bar{\mu}_{2s;R}$ , ( $s = 0, 1, \dots, 5$ ) are given by (2.10).

If the class marks are selected as the unit of  $x$ , we set  $h = 1$  in (2.10). If the

class interval is chosen as the unit of  $x$ , we set  $h = 1/k$  in (2.10). If  $k$  consecutive values of the discrete variable are grouped in a frequency class of width  $m$ , we put  $h = m/k$  in (2.10).

Usually we desire to estimate the value of the moments that would have been obtained if we had not grouped the data. Therefore (2.13) is solved for the moments of the ungrouped data. We have

$$(2.14) \quad \mu_t = \sum_{s=0}^{\lfloor t/2 \rfloor} \binom{t}{2s} P_{2s} \mu_{t-2s}$$

wherein

$$P_{2s} = \sum \frac{(-1)^s (2s)! \rho! \bar{\mu}_{2p_1:R}^{\pi_1} \bar{\mu}_{2p_2:R}^{\pi_2} \cdots \bar{\mu}_{2p_v:R}^{\pi_v}}{[(2p_1)!]^{\pi_1} [(2p_2)!]^{\pi_2} \cdots [(2p_v)!]^{\pi_v} \pi_1! \cdots \pi_v!},$$

the summation being taken for every possible product of moments for which

$$\sum_{i=1}^v p_i = s, \quad \sum_{i=1}^v \pi_i = \rho.$$

Formulas, corresponding to (2.13) and (2.14), for a distribution of continuous variates are written by replacing the moment symbols for discrete variates by those for continuous variates.

**3. Adjustments for central moments.** Consider the universe  $U$  which consists of the population  $,X$ , ( $r = 1, 2, \dots, k$ ), where  $,X$  is the  $r$ th grouped-distribution.

The expected value of the  $t$ th central moment of the  $k$  grouped-distribution is given by (1.3), (1.4) and (1.5) of Part I, where now  $\mu_{1:\mu_t-i}$  is given by (2.13) of the preceding section. Thus, the development of this section is identical with that of section one of Part I with the single exception that  $\mu_{1:\mu_t} = \mu_t$  no longer holds but is replaced by  $\mu_{1:\mu_t} = \mu_t + \text{a correction}$ . Therefore, the formulas for the adjustments for central moments may be obtained immediately from the formulas derived in section one, Part I, if the corrections of the preceding section are inserted. We have

$$(2.15) \quad \mu_{1:\bar{\mu}_2} = \bar{\mu}_2 + \bar{\mu}_{2:R} - \bar{\mu}_{2:\mu_1}.$$

$$(2.16) \quad \mu_{1:\bar{\mu}_3} = \bar{\mu}_3 + 6\mu_1\bar{\mu}_{2:\mu_1} - 3\bar{\mu}_{11:\mu_1\mu_2} + 2\bar{\mu}_{3:\mu_1}.$$

$$(2.17) \quad \begin{aligned} \mu_{1:\bar{\mu}_4} = & \bar{\mu}_4 + 6\bar{\mu}_2\bar{\mu}_{2:R} + \bar{\mu}_{4:R} + 6(\bar{\mu}_2 - 2\mu_1^2 + \bar{\mu}_{2:R})\bar{\mu}_{2:\mu_1} \\ & + 12\mu_1\bar{\mu}_{11:\mu_1\mu_2} - 12\mu_1\bar{\mu}_{3:\mu_1} - 4\bar{\mu}_{11:\mu_1\mu_2} \\ & + 6\bar{\mu}_{21:\mu_1\mu_2} - 3\bar{\mu}_{4:\mu_1}. \end{aligned}$$

The moments of the ungrouped data can be obtained readily from formulas (2.15) through (2.17).

Adjustment formulas for central moments of a distribution of continuous variates may be obtained from (2.13) by replacing the moment symbols for

discrete variates by those for continuous variates and taking the moments about the mean. Also, it may be observed that adjustment formulas for central moments of a distribution of continuous variates may be obtained from formulas (1.3), (1.4) and (1.5) of Part I, provided the moment symbols are exchanged as indicated above and terms of the form  $\bar{\nu}_{r_1 r_2 \dots r_t; \nu_{t_1} \nu_{t_2} \dots \nu_{t_r}}$  are set equal to zero.

**4. Usual adjustments for Thiele seminvariants.** The usual adjustments for Thiele seminvariants, for the univariate discrete population, may be developed directly by use of one of the fundamental properties of Thiele seminvariants.

It is assumed (see [25; p. 55]) that  $k$  consecutive values of the discrete variable are grouped in a frequency class of width  $m$ . The  $k$  smaller intervals of width  $m/k = h$  go to make up the class width  $m$ , the actual points representing the  $k$  values of the variable being plotted at the centers of the sub-intervals. Now, let us suppose that each of the  $k$  consecutive boundary points of the subintervals is as likely to be chosen as a boundary point of the larger intervals as any other. Then, if  $x_i$  is the class mark of the  $i$ th frequency class, for any true value,  $x$ , of the discrete variable included in this frequency class, we have

$$x_i = x + e_r$$

in which  $x$  and  $e_r$  are independent variables and  $e_r$  takes on the  $k$  values (2.12) with equal relative frequencies  $1/k$ .

Since we have noted that the equally likely values which  $e_r$  may take on are deviations about the mean of a rectangular distribution of discrete variates, we employ the cumulative property of Thiele seminvariants [9; p. 4] and obtain directly

$$(2.18) \quad \lambda'_{t;x} = \lambda_{t;x} + \lambda_{t;R}, \quad (t = 1, 2, \dots),$$

where  $\lambda'_{t;x}$  is the  $t$ th seminvariant computed from the grouped data,  $\lambda_{t;x}$  is the  $t$ th seminvariant computed from the ungrouped data and  $\lambda_{t;R}$  is defined by (2.11).

Formulas corresponding to (2.18), for special values of  $t$ , are given by Craig [25; p. 57]. However, the present development indicates the dependence of adjustment formulas on central moments of a rectangular distribution and provides a general formula for these adjustments which is expressed completely in terms of Thiele seminvariants.

**5. New adjustments for Thiele seminvariants.** If we accept the definition

$$\mu_{1;\hat{\mu}_t} = \frac{1}{k} \sum_{r=1}^k r \bar{\mu}_t, \quad (t = 2, 3, \dots),$$

then (2.18) is at best only an approximation formula. We now desire exact formulas for  $\mu_{1;\lambda_t}$  for the case of a grouped-distribution of discrete variates.

First (1.9) is used and terms of the form  $\mu_{s_1 s_2 \dots s_r; \mu_1 \mu_2 \dots \mu_v}$  are evaluated in terms of central moments by (1.3). Then terms of the form  $\mu_{1; \mu_t}$  are evaluated by (2.13) and finally the relations between moments and Thiele seminvariants are employed. Exact formulas for the expected values of the second, third, and fourth Thiele seminvariants for grouped-distributions of discrete variables are given below.

$$(2.19) \quad \mu_{1; \lambda_2} = \lambda_2 + \lambda_{2; R} - \bar{\mu}_{2; \mu_1}.$$

$$(2.20) \quad \mu_{1; \lambda_3} = \lambda_3 + 6\lambda_1 \bar{\mu}_{2; \mu_1} - 3\bar{\mu}_{11; \mu_1 \mu_2} + 2\bar{\mu}_{3; \mu_1}.$$

$$(2.21) \quad \begin{aligned} \mu_{1; \lambda_4} = & \lambda_4 + \lambda_{4; R} + 12[\lambda_2 - 2\lambda_1^2 + \lambda_{2; R}]\bar{\mu}_{2; \mu_1} \\ & + 24[\bar{\mu}_{11; \mu_1 \mu_2} - \bar{\mu}_{3; \mu_1}]\lambda_1 - 4\bar{\mu}_{11; \mu_1 \mu_2} \\ & + 12\bar{\mu}_{21; \mu_1 \mu_2} - 6\bar{\mu}_{4; \mu_1} - 3\bar{\mu}_{2; \mu_2}. \end{aligned}$$

Formulas for Thiele seminvariants of ungrouped data in terms of expectations may be obtained from (2.19) through (2.21).

Adjustment formulas for Thiele seminvariants of a distribution of continuous variates are given by Langdon and Ore [23; p. 231] and Craig [25; p. 57]. If we denote the  $t$ th Thiele seminvariant of a distribution of continuous variates by  $L_t$ , then

$$(2.22) \quad \nu_{1; L_t} = L_t + L_{t; R},$$

where

$$(2.23) \quad L_{2t+1; R} = 0, \quad L_{2t; R} = \frac{(-1)^{t-1} B_t m^{2t}}{2t}, \quad t = 1, 2, \dots$$

Formulas (2.19) through (2.21) may be used for continuous variates by changing the moment symbols and setting terms of the form  $\bar{\mu}_{s_1 s_2 \dots s_r; \mu_{t_1} \mu_{t_2} \dots \mu_{t_r}}$  equal to zero.

**6. Adjustment formulas applied to a numerical problem.** We consider the arbitrary distribution given in Table III.

TABLE III  
*An Arbitrary Distribution of Discrete Variates*

$v$	$f$	$v$	$f$	$v$	$f$	$F$
1	2	4	30	7	1	$2 + 30 + 1 = 33$
2	8	5	4	8	1	$8 + 4 + 1 = 13$
3	10	6	3	9	1	$10 + 3 + 1 = 14$



The three grouped distributions, when the variates are grouped in "groupings of three," appear in Table IV.

TABLE IV

*Distributions Derived from Data of Table III by Making the Three Possible Groupings of Three*

(1)		(2)		(3)	
Class	$f$	Class	$f$	Class	$f$
1-3	20	0-2	10	-1 to 1	2
4-6	37	3-5	44	2-4	48
7-9	3	6-8	5	5-7	8
10-12	0	9-11	1	8-10	2

Using the fixed point 4, moment-functions are computed for the distribution of Table III and for each of the distributions of Table IV. These quantities along with the average of each moment function appear in Table V.

TABLE V

*Moment-Functions of the Distributions of Table III and Table IV. Averages of Moment-Functions of Distributions of Table IV*

Dist.	$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$	$\bar{\mu}_2 = \lambda_2$	$\bar{\mu}_3 = \lambda_3$	$\bar{\mu}_4$	$\lambda_4$
(1)	$\frac{9}{60}$	$\frac{165}{60}$	$\frac{69}{60}$	$\frac{1125}{60}$	$\frac{9819}{(60)^2}$	$\frac{-17442}{(60)^3}$	$\frac{238,849,317}{(60)^4}$	$\frac{-50,388,966}{(60)^4}$
(2)	$\frac{-9}{60}$	$\frac{171}{60}$	$\frac{81}{60}$	$\frac{2511}{60}$	$\frac{10179}{(60)^2}$	$\frac{567162}{(60)^3}$	$\frac{557,840,277}{(60)^4}$	$\frac{247,004,154}{(60)^4}$
(3)	$\frac{-30}{60}$	$\frac{162}{60}$	$\frac{138}{60}$	$\frac{1938}{60}$	$\frac{8820}{(60)^2}$	$\frac{1317600}{(60)^3}$	$\frac{528,282,000}{(60)^4}$	$\frac{294,904,800}{(60)^4}$
Ave.	$\frac{-10}{60}$	$\frac{166}{60}$	$\frac{96}{60}$	$\frac{1858}{60}$	$\frac{9606}{(60)^2}$	$\frac{622440}{(60)^3}$	$\frac{441,657,198}{(60)^4}$	$\frac{163,839,996}{(60)^4}$
Orig. Dist.	$\frac{-10}{60}$	$\frac{126}{60}$	$\frac{116}{60}$	$\frac{1314}{60}$	$\frac{7460}{(60)^2}$	$\frac{642400}{(60)^3}$	$\frac{305,034,000}{(60)^4}$	$\frac{138,079,200}{(60)^4}$

Table VI gives the expected values of the moment-functions as obtained by substituting from Table V into the formulas of sections two, three, and five. Also the expected values, computed from the usual formulas, are given and the errors which would be made, if the usual formulas were used, are indicated.

TABLE VI  
*Expected Values of Moment-Functions Computed by Formulas*

Expectations by	$\mu_{1:\mu_1}$	$\mu_{1:\mu_2}$	$\mu_{1:\mu_3}$	$\mu_{1:\mu_4}$	$\mu_{1:\bar{\mu}_2} = \mu_{1:\lambda_2}$	$\mu_{1:\bar{\mu}_3} = \mu_{1:\lambda_3}$	$\mu_{1:\bar{\mu}_4}$	$\mu_{1:\lambda_4}$
New Formulas	$\frac{-10}{60}$	$\frac{166}{60}$	$\frac{96}{60}$	$\frac{1858}{60}$	$\frac{9806}{(60)^2}$	$\frac{622440}{(60)^3}$	$\frac{441,657,198}{(60)^4}$	$\frac{163,839,996}{(60)^4}$
Usual Formulas	$\frac{-10}{60}$	$\frac{166}{60}$	$\frac{96}{60}$	$\frac{1858}{60}$	$\frac{9860}{(60)^2}$	$\frac{642400}{(60)^3}$	$\frac{416,778,000}{(60)^4}$	$\frac{133,795,200}{(60)^4}$
Error	—	—	—	—	$\frac{254}{(60)^2}$	$\frac{19960}{(60)^3}$	$\frac{-24,879,198}{(60)^4}$	$\frac{-30,060,796}{(60)^4}$

**7. Evaluation of  $\bar{\mu}_{2:\mu_1}$ .** It appears at first that it is necessary to form the " $k$  groupings of  $k$ " in order to evaluate the term  $\bar{\mu}_{2:\mu_1}$  which enters the precise formula for the expected value of the variance. That was the procedure followed by Carver [26; p. 161]. However, it is possible to evaluate  $\bar{\mu}_{2:\mu_1}$  from the ungrouped data without forming a single grouped-distribution.

By definition,

$$\bar{\mu}_{2:\mu_1} = \frac{1}{k} \sum_{r=1}^k [\tau\mu_1 - \mu_1]^2,$$

where  $\tau\mu_1$  is the mean of the  $\tau$ th grouped-distribution and  $\mu_1$  is the mean of the ungrouped distribution. We wish to study the terms  $\tau\mu_1$  and  $\mu_1$ . Consider a set of variates  $x_i$ , ( $i = 1, 2, \dots, s$ ), with corresponding frequencies  $f_i$ , ( $i = 1, 2, \dots, s$ ). The  $x$ 's are subject to the condition,  $x_i - x_{i-1} = 1$ , and consequently some of the  $f$ 's may be zero. The mean of this distribution is  $\frac{\sum xf}{\sum f}$ .

We define

$$F_i = f_i + f_{k+i} + f_{2k+i} + \dots, \quad (i = 1, 2, \dots, k)$$

Then, if a grouped-distribution is formed with  $x_i$  in the  $i$ th ( $i = 1, 2, \dots, k$ ) position of a class, the mean of this grouped-distribution is

$$\frac{\sum xf + \sum_{j=1}^k F_j e_{i+j-1}}{\sum f}$$

where  $e_{i-1} = e_k$  if  $e_i = 1$  and  $e_{i+1} = e_1$  if  $e_i = e_k$ . Similarly if a grouped-distribution is formed with  $x_i$  in the  $(i+1)$ st position of a class, the mean is

$$\frac{\sum xf + \sum_{j=1}^k F_j e_{i+j}}{\sum f}.$$

Thus, it is evident that, given the expression for the mean of any grouped-distribution in which  $x_i$  is in the  $i$ th position of a class, we may form the expression for the mean of the grouped-distribution in which  $x_i$  is in the  $(i + 1)$ st position of a class by a cyclic permutation of the  $e_i$ 's of the given expression.

Therefore, it follows that if we call  ${}_{r\mu_1}$  the mean of the grouped-distribution in which  $x_i$  is in the  $r$ th ( $r = 1, 2, \dots, k$ ) position of a class, then

$${}_{r\mu_1} - \mu_1 = \frac{\sum_{j=1}^k F_j e_{r+j-1}}{\sum f}, \quad (r = 1, 2, \dots, k).$$

If we define

$$N = \sum f \quad \text{and} \quad \phi_r = \sum_{j=1}^k F_j e_{r+j-1}$$

then,

$$\bar{\mu}_{2;\mu_1} = \frac{1}{kN^2} \sum_{r=1}^k \phi_r^2.$$

Thus, it is evident that  $\bar{\mu}_{2;\mu_1}$  is a function of the frequencies of the variates and of the  $e_i$ 's. The fact that the values of the variates do not enter  $\bar{\mu}_{2;\mu_1}$  permits one to quickly calculate its value.

Consider  $\bar{\mu}_{2;\mu_1}$  for the distribution of Table III. We find

$$\phi_1 = 33e_1 + 13e_2 + 14e_3.$$

Then, by successive cyclic permutations of the  $e_i$ 's,

$$\phi_2 = 33e_2 + 13e_3 + 14e_1,$$

$$\phi_3 = 33e_3 + 13e_1 + 14e_2.$$

Substituting the values  $e_1 = 1$ ,  $e_2 = 0$ ,  $e_3 = -1$  we have  $\phi_1 = 19$ ,  $\phi_2 = 1$  and  $\phi_3 = -20$ . Therefore,

$$\bar{\mu}_{2;\mu_1} = \frac{254}{(60)^2}$$

which is identical with the value which was found when Table V was used.

It follows from the preceding development that

$$\bar{\mu}_{t;\mu_1} = \frac{1}{kN^t} \sum_{r=1}^k \phi_r^t$$

and if  $F_1 = F_2 = \dots = F_k$  then  $\bar{\mu}_{t;\mu_1}$  is zero.

**8. Conclusion.** The results of this paper include:

1. The derivation of general and specific formulas for the expected values of population moment-functions.

2. The derivation of generalized sampling formulas under the condition that samples of  $n$  are formed by selecting one variate from each population.
3. Methods for the transformation of generalized sampling formulas into the corresponding infinite and finite sampling formulas.
4. A method for the transformation of infinite sampling formulas into the corresponding finite sampling formulas.
5. A demonstration of the fact that adjustment formulas for moment-function of grouped data involve central moments of a rectangular distribution.
6. A general formula for the expected value of the  $t$ th moment of grouped data.
7. New adjustment formulas for central moments of grouped data.
8. New adjustment formulas for Thiele seminvariants of grouped data.
9. A method for the evaluation of the term  $\bar{\mu}_{2;\mu_1}$  which appears in the precise adjustment formula for the variance.

Many thanks are due Prof. P. S. Dwyer, to whom the writer is greatly indebted for advice and encouragement.

#### REFERENCES

- [1] "STUDENT", "The Probable Error of the Mean," *Biometrika*, Vol. 6 (1908), pp. 1-25.
- [2] L. ISSERLIS, "On the Value of a Mean as Calculated from a Sample," *Jour. of Roy. Stat. Soc.*, Vol. 81 (1918), pp. 75-81.
- [3] A. TCHOUPROFF, "On the Mathematical Expectation of the Moments of Frequency Distributions," Part I, *Biometrika*, Vol. 12 (1918), pp. 140-169, 185-210; Part II, *Biometrika*, Vol. 13 (1920-21), pp. 283-295.
- [4] A. TCHOUPROFF, "On the Mathematical Expectation of the Moments of Frequency Distributions in the Case of Correlated Observations," *Metron*, Vol. 2 (1923), pp. 461-493; 646-683.
- [5] A. CHURCH, "On the Moments of the Distributions of Squared Deviations for Samples of  $N$  Drawn from an Indefinitely Large Population," *Biometrika*, Vol. 17 (1925), pp. 79-83.
- [6] J. NEYMAN, "Contributions to the Theory of Small Samples Drawn from a Finite Population," *Biometrika*, Vol. 17 (1925), pp. 472-479.
- [7] A. CHURCH, "On the Means and Squared Deviations of Small Samples from any Population," *Biometrika*, Vol. 18 (1926), pp. 321-394.
- [8] K. PEARSON, "Another Historical Note on the Problem of Small Samples," *Biometrika* Vol. 19 (1927), pp. 207-210.
- [9] C. C. CRAIG, "An Application of Thiele's Seminvariants to the Sampling Problem," *Metron*, Vol. 7 (1928-29), pp. 3-74.
- [10] R. A. FISHER, "Moments and Product Moments of Sampling Distributions," *Proc. London Math. Soc.*, Vol. 2 (30) (1929), pp. 199-238.
- [11] J. PEPPER, "Studies in the Theory of Sampling," *Biometrika*, Vol. 11 (1929), pp. 231-258.
- [12] H. C. CARVER, "Fundamentals of Sampling," *Annals of Math. Stat.*, Vol. 1 (1930), pp. 101-121, 260-274.
- [13] N. ST. GEORGESQUE, "Further Contributions to the Sampling Problem," *Biometrika*, Vol. 24 (1932), pp. 65-107.
- [14] G. M. BROWN, "On Sampling from Compound Populations," *Annals of Math. Stat.*, Vol. 4 (1933), pp. 288-342.
- [15] J. NEYMAN, "On the Two Different Aspects of the Representative Method," *Jour. Roy. Stat. Soc.*, Vol. 97 (1934), pp. 558-625.

- [16] A. L. BOWLEY, "Measurement of the Precision Attained in Sampling," *Bull. International Stat. Inst.*, Vol. 22.
- [17] H. FELDMAN, "Mathematical Expectations of Product Moments of Samples Drawn from a Set of Infinite Populations," *Annals of Math. Stat.*, Vol. 6 (1935), pp. 30-52.
- [18] H. C. CARVER, "Sheppard's Corrections for a Discrete Variable," *Annals of Math. Stat.*, Vol. 76 (1936), pp. 154-163.
- [19] P. S. DWYER, "Moments of Any Rational, Integral Isobaric Sample Moment Function," *Annals of Math. Stat.*, Vol. 8 (1937), pp. 21-65.
- [20] P. S. DWYER, "Combined Expansions of Products of Symmetric Power Sums and of Sums of Symmetric Power Products with Application to Sampling," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 1-84.
- [21] A. T. CRAIG, "On the Mathematics of the Representative Method of Sampling," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 26-34.
- [22] A. FISHER, *The Mathematical Theory of Probabilities*, New York, The Macmillan Company. 1928.
- [23] W. H. LANGDON, and ORE, O., "Seminvariants and Sheppard's Corrections," *Annals of Math.*, Vol. 31 (1930), pp. 230-232.
- [24] J. R. ABERNETHY, "On the Elimination of Systematic Errors Due to Grouping," *Annals of Math. Stat.*, Vol. 4 (1933), pp. 263-277.
- [25] C. C. CRAIG, "Sheppard's Corrections for a Discrete Variable," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 55-61.
- [26] H. C. CARVER, "The Fundamental Nature and Proof of Sheppard's Adjustments," *Annals of Math. Stat.*, Vol. 9 (1936), pp. 154-163.
- [27] M. SASULY, *Trend Analysis of Statistics*, Washington, The Brookings Institution (1934).

THE UNIVERSITY OF MICHIGAN  
ANN ARBOR, MICHIGAN

# THE ANALYSIS OF VARIANCE WHEN EXPERIMENTAL ERRORS FOLLOW THE POISSON OR BINOMIAL LAWS

BY W. G. COCHRAN

**1. Introduction.** The use of transformations has recently been discussed by several writers [1], [2], [3], [4], in applying the analysis of variance to experimental data where there is reason to suspect that the experimental errors are not normally distributed. Two types of transformations appear to be coming into fairly common use:  $\sqrt{x}$  and  $\sin^{-1} \sqrt{x}$ . The former is considered appropriate where the data are small integers whose experimental errors follow the Poisson law, while the latter applies to fractions or percentages derived from the ratio of two small integers, where the experimental errors follow the binomial frequency distribution. In each case the object of the transformation is to put the data on a scale in which the experimental variance is approximately the same on all plots, so that all plots may be used in estimating the standard error of any treatment comparison. The extent to which these transformations are likely to succeed in so doing has been examined by Bartlett [2]. The object of the present paper is to discuss the theoretical basis for these transformations in more detail, and in particular to examine their relation to a more exact analysis.

**2. Experimental variation of the Poisson type.** The first step in an exact statistical analysis of the results of any field experiment, is to specify in mathematical terms (1) how the expected values on each plot are obtained in terms of unknown parameters representing the treatment and block (or row and column) effects (2) how the observed values on the plots vary about the expected values. In this section, the variation is assumed to follow the Poisson law.

The specification of the expected values requires some consideration. In the standard theory of the analysis of variance, treatment and block (or row and column) effects are assumed to be additive. In the case of a Latin square, for example, the expected yield  $m_{it}$  of the  $i$ th plot, which receives the  $t$ th treatment and occurs in the  $r$ th row and the  $c$ th column is written

$$(1) \quad m_{it} = G + T_t + R_r + C_c$$

where  $G$  is a parameter representing the average level of yield in the experiment, and  $T_t$ ,  $R_r$ , and  $C_c$  represent the respective effects of the treatment, row and column to which the plot corresponds. Since the  $T$ ,  $R$  and  $C$  constants are required only to measure differences between different treatments, rows and columns, we may put

$$(2) \quad \sum_i T_i = \sum_r R_r = \sum_c C_c = 0.$$

If the experimental errors are normally and independently distributed with equal variance, this specification leads to very simple equations of estimation for the unknown parameters, the maximum likelihood estimate of  $T_i$ , for example, being the difference between the mean yield of all plots receiving that treatment and the general mean. In addition to its simplicity, this type of prediction formula is fairly suitable for general use, because it gives a good approximation to most types of law which might be envisaged, provided that row and column differences are small in relation to the mean yield. However, in considering an exact analysis with Poisson variation, the prediction formula is assumed chosen, without reference to computational simplicity, as being the most suitable to describe the combined actions of treatment and soil effects.

The probability of obtaining a given set of plot yields  $x_i$  with expectations  $m_i$  may be written

$$\prod_i \frac{e^{-m_i} m_i^{x_i}}{x_i!}.$$

Thus  $L$ , the logarithm of the likelihood, is given by

$$(3) \quad L = \sum_i (x_i \log m_i - m_i) - \sum_i \log x_i!.$$

Hence the maximum likelihood equation of estimation for any parameter  $\theta$  assumes the form

$$(4) \quad \sum \frac{(x_i - m_i)}{m_i} \frac{\partial m_i}{\partial \theta} = 0$$

where the summation extends over all plots whose expectations involve  $\theta$ . The function  $\frac{\partial m_i}{\partial \theta}$  will usually involve a number of parameters. Since the specification of row, column and treatment effects in a 6 x 6 Latin square requires 16 independent parameters, the solution of these equations may be expected to be laborious, though it may be shortened by the intelligent use of iterative methods. The problem of obtaining exact tests of significance is also difficult. The method of maximum likelihood provides estimates of the variances and co-variances of the treatment constants, which under certain conditions can be assumed to be normally distributed if there is sufficient replication, but this can hardly be considered an exact "small sample" solution.

These remarks show that the exact solution is somewhat too complicated for frequent use. The difficulty arises principally because the typical equation of estimation consists of a *weighted* sum of the deviations of the observed from the expected values, the weights being  $\frac{1}{m_i} \frac{\partial m_i}{\partial \theta}$ . The factor  $\frac{1}{m_i}$  was introduced into the weight by the Poisson variation of the experimental errors, and must be retained in any theory which claims to apply to Poisson variation. It is, however, worth considering whether some simplification cannot be introduced into

the equations by assuming some particular form for the prediction formula. This line of approach seems promising when one considers the simplification introduced into the "normal theory" case by assuming the prediction formula to be linear.

For Poisson variation, the linear law does not appear to be particularly suitable, since it may give negative expectations on some plots (as happens in the numerical example considered in the next section). Further, while  $\frac{\partial m_i}{\partial \theta}$  becomes a constant, the factor  $\frac{1}{m_i}$  remains in the weight.

The entire weight can be made constant by assuming a linear prediction formula in the square roots and transforming the data to square roots. For a Latin square, this prediction formula is written

$$(5) \quad \sqrt{m_i} = \alpha_i = G + T_t + R_r + C_c,$$

where

$$(6) \quad \sum_t T_t = \sum_r R_r = \sum_c C_c = 0.$$

To find the maximum value of (3) subject to the restrictions (6), we may use the method of undetermined multipliers, maximizing

$$(7) \quad L + \lambda(\sum_t T_t) + \mu(\sum_r R_r) + \nu(\sum_c C_c).$$

The equation of estimation for a typical treatment constant  $T_t$  becomes

$$(8) \quad \sum \left( \frac{x_i - m_i}{m_i} \right) \frac{dm_i}{d\alpha_i} \frac{\partial \alpha_i}{\partial T_t} + \lambda = 0, \quad \text{i.e.,} \quad \sum \frac{2(x_i - m_i)}{\sqrt{m_i}} + \lambda = 0,$$

the summation being extended over all plots receiving the treatment. If  $a_i = \sqrt{x_i}$ , then by Taylor's theorem

$$(9) \quad x_i - m_i = (a_i - \alpha_i) \frac{dm_i}{d\alpha_i} + \frac{1}{2!} (a_i - \alpha_i)^2 \frac{d^2 m_i}{d\alpha_i^2} + \dots$$

If  $m_i$  is reasonably large, only the first term on the right-hand side need be retained. When  $m_i$  is small, we may use, instead of the exact square root, a quantity  $a'_i$  defined so that

$$(10) \quad x_i - m_i = (a'_i - \alpha_i) \frac{dm_i}{d\alpha_i} = 2\sqrt{m_i}(a'_i - \alpha_i).$$

Thus if the analysis is performed on the quantities  $a'_i$  instead of on the original data, equation (8) becomes

$$(11) \quad \sum_{T_t} 4(a'_i - \alpha_i) + \lambda = 0.$$



On substituting the expectations for  $\alpha_i$  from (5), and using (6), we obtain

$$(12) \quad \sum_i 4(a'_i - G - T_i) + \lambda = 0.$$

The corresponding equation for  $G$  is

$$(13) \quad \sum_i 4(a'_i - G) = 0,$$

so that  $G$  is the general mean of the quantities  $a'$ . By adding equations (12) over all treatments, and comparing the total with (13), we find  $\lambda = 0$ . Hence  $T_i$  is the difference between the mean yield of  $a'$  over all plots receiving  $T_i$  and the general mean of  $a'$ . In this scale the simplicity of the "normal theory" equations has apparently been recovered. Actually, the quantities  $a'$  are not known exactly, since

$$(14) \quad a' = \alpha + \frac{(x - m)}{2\sqrt{m}} = \frac{1}{2} \left( \alpha + \frac{x}{\alpha} \right)$$

where  $\alpha$  is the expected value of  $\sqrt{x}$ . However, this process provides a means of successively approximating the maximum likelihood solution, by choosing first approximations to the quantities  $\alpha$ , constructing the  $a'$ 's, solving for the unknown constants and hence obtaining second approximations to the expected values. The close relation of  $a'$  to  $\sqrt{x}$  is seen by remembering one of the common rules for finding square roots. This consists in guessing an approximate root ( $\alpha$ ), dividing  $x$  by the approximate root, and taking the mean of the approximate root ( $\alpha$ ) and the resulting quotient ( $x/\alpha$ ).

The suitability of the linear prediction formula in square roots must be considered in any example in which the above analysis is being employed. The law is intermediate in its effects between the linear law and the product law in the original data. My experience is that it is fairly satisfactory for general use, (cf. [2], p. 72). An exception may occur when it is desired to test the interaction between two treatments, both of which produce large effects. In this case the definition chosen for absence of interaction may not coincide at all closely with the definition implied in using the linear law in square roots. An example of this case was given in a previous paper [1].

In this connection it should be noted that an approximate "goodness of fit" test may be obtained of the validity of the assumptions made. Since the quantities  $a'_i$  enter into the equations of estimation with weight 4, the quantity  $4 \sum_i (a'_i - \alpha_i)^2$  is distributed approximately as  $\chi^2$  with the number of degrees of freedom in the error term of the analysis of variance. Some idea of the closeness of the approximation may be gathered by considering the simplest case in which only the mean yield is being estimated. In this case the observed values  $x$  are assumed to be drawn from the same Poisson distribution, and the sufficient statistic for the mean  $G$  is known to be  $\Sigma(x_i)/n$ . Since, however, the

prediction formula is here the same in square roots as in the original scale, and since the maximum likelihood solution is invariant to change of scale, the mean value  $\alpha$  of  $a'$  must be *exactly*  $\sqrt{\Sigma(x)/n}$ , as the reader may verify by working any particular example. Thus  $\Sigma 4(a' - \alpha)^2$  is found to be  $\Sigma(x - \bar{x})^2/\bar{x}$ , the usual  $\chi^2$  test for examining whether a set of values  $x$  may reasonably be assumed to come from the same Poisson distribution. By working out the exact distribution of  $\Sigma(x_i - \bar{x})^2/\bar{x}$  in a number of cases [5], I previously expressed the opinion that this quantity followed the  $\chi^2$  distribution sufficiently closely for most practical uses, even for values of the mean as low as 2. This opinion has since been substantiated by Sukhatme, [6] who sampled this distribution for  $m = 1, 2, 3, 4$ , and 5.

A high value of  $\chi^2$  means either that the prediction formula is not satisfactory or that the experimental errors are higher than the Poisson distribution indicates, or that both causes are operating. These effects can sometimes be separated by examining whether the observed yields deviate from the expected yields in a systematic or a random manner. If the deviation is systematic, the prediction formula is probably unsatisfactory.

The type of approach used above resembles in many features the "exact" analysis for the probit transformation [7]. The principal difference is that in the case of probits the transformation is made to suit the *a priori* prediction formula, which postulates that the probits are a linear function of the dosage, or of the log (dosage). Thus with probits the equations of estimation still involve weights in the transformed scale. These do not seriously complicate the analysis, since only two parameters require to be estimated for a given poison. With, however, the much greater number of parameters usually involved in specifying the results of a field experiment, the attractiveness of a solution which does not involve weighting is greatly increased.

**3. Numerical example of the square root transformation.** A  $5 \times 5$  Latin square experiment on the effects of different soil fumigants in controlling wireworms was selected as an example. The average number of wireworms per plot (total of four soil samples) was just under five. Previous studies [8], [9] have indicated that with small numbers per sample, the distribution of numbers of wireworms tends to follow the Poisson law.

The plan and yields are shown in Table I. The first two figures under the treatment symbols are the numbers of wireworms and their square roots respectively, the latter being regarded as first approximations to the values  $a'$ . Two of the plots receiving treatment *K* gave no wireworms. Since these plots are likely to be changed most in the transition from square roots to  $a'$ , better approximations were estimated for them before proceeding with the calculations. The best simple approximations appeared to be obtained from the square roots of the means in the original units. For the plot in the second row and second column, the square roots of the row, column and treatment means in the original

TABLE I

*Plan and number of wireworms per plot*

<i>P</i>	<i>O</i>	<i>N</i>	<i>K</i>	<i>M</i>	Mean
3 <sup>1</sup>	2	5	1	4	
1.73 <sup>2</sup>	1.41	2.24	1.00	2.00	1.676 <sup>2</sup>
1.76 <sup>3</sup>	1.45	2.25	1.11	2.00	1.714 <sup>3</sup>
1.77 <sup>4</sup>	1.46	2.25	1.10	2.00	1.716 <sup>4</sup>
<i>M</i>	<i>K</i>	<i>O</i>	<i>N</i>	<i>P</i>	
6	0	6	4	4	
2.45	(0.39)	2.45	2.00	2.00	1.858
2.45	0.32	2.50	2.02	2.02	1.862
2.46	0.32	2.49	2.02	2.02	1.862
<i>O</i>	<i>M</i>	<i>K</i>	<i>P</i>	<i>N</i>	
4	9	1	6	5	
2.00	3.00	1.00	2.45	2.24	2.138
2.10	3.09	1.00	2.47	2.25	2.182
2.13	3.08	1.00	2.46	2.25	2.184
<i>N</i>	<i>P</i>	<i>M</i>	<i>O</i>	<i>K</i>	
17	8	8	9	0	
4.12	2.83	2.83	3.00	(0.79)	2.714
4.18	2.84	2.83	3.00	0.77	2.724
4.17	2.84	2.83	3.00	0.77	2.722
<i>K</i>	<i>N</i>	<i>P</i>	<i>M</i>	<i>O</i>	
4	4	2	4	8	
2.00	2.00	1.41	2.00	2.83	2.048
2.14	2.02	1.49	2.04	2.92	2.122
2.10	2.03	1.50	2.05	2.90	2.116
<i>Mean</i>	2.460 <sup>2</sup>	1.926	1.986	2.090	1.972
	2.526 <sup>3</sup>	1.944	2.014	2.128	1.992
	2.526 <sup>4</sup>	1.946	2.014	2.126	1.988

## Treatment Means

<i>K</i>	<i>P</i>	<i>O</i>	<i>M</i>	<i>N</i>
1.036 <sup>2</sup>	2.084	2.338	2.456	2.520
1.068 <sup>3</sup>	2.116	2.394	2.482	2.544
1.058 <sup>4</sup>	2.118	2.396	2.484	2.544

<sup>1</sup>Original numbers. <sup>2</sup>Square roots. <sup>3</sup>Second approximations. <sup>4</sup>Third approximations.

units are respectively 2.000, 2.145 and 1.095, and the square root of the general mean is 2.227. Hence

$$a' = \frac{1}{3}[2.000 + 2.145 + 1.095 - 2(2.227)] = 0.39.$$

The other zero value was similarly found to give  $a' = 0.79$ . The corresponding estimates from the means of the square roots were considerably too low, since the  $a'$  values tend to be higher than the square roots. The use of "missing plot" technique gave very poor approximations, because it ignores the fact that the plots in question had zero yields.

With the estimated values inserted, the row, column, and treatment means of the square roots are as shown in Table I. A second approximation to  $a'$  was calculated for each plot. For the plot in the first row and the first column, the expected yield is

$$\alpha = 1.676 + 2.460 + 2.084 - 2(2.087) = 2.046.$$

Hence  $a' = \frac{1}{3}(2.046 + 3/2.046) = 1.76$ . These values constitute the third set of figures in Table I. Theoretically, it is advisable to readjust the row, column, and treatment means after each new value of  $a'$  has been obtained, in order to secure rapid convergence. This is rather laborious in practice, and a complete set of new plot values was obtained before readjusting the means. The third approximations obtained by this method are shown in the fourth lines in Table I and are correct to two decimal places.

It is noteworthy how closely the square roots agree with the third approximations on all plots except those which originally gave zero yields. The differences between the second and third approximations are trivial.

The next step is to make a  $\chi^2$  test by means of the quantity  $4\Sigma(a' - \alpha)^2$ . From the manner in which the values  $\alpha$  are constructed from the  $a'$ 's, it follows that  $\Sigma(a' - \alpha)^2$  is simply the error sum of squares in the conventional analysis of variance of the values  $a'$ . The analysis of variance of the third approximations is shown in Table II.

TABLE II  
*Analysis of variance of adjusted square roots*

	Degrees of freedom	Sum of squares	Mean square
Rows	4	2.9815	
Columns	4	1.1190	
Treatments	4	7.5815	1.8954
Error	12	4.5970	0.3831

The value of  $\chi^2$  is  $4 \times 4.597 = 18.39$ , with 12 degrees of freedom, which is just about the 10 percent level. If the hypothesis is regarded as disproved only when  $\chi^2$  exceeds the 5 percent level, the treatment means may be tested by regarding them as approximately normally distributed with variance

$1/5 \times 0.25 = 0.05$ . It is, however, more prudent to use the actual error mean square as an estimate of the experimental error variance, performing the usual tests associated with the analysis of variance. This may be justified on the grounds that the calculations have produced a set of plot values  $a'$  of equal weight. On this basis the standard error of a treatment mean is  $\sqrt{0.3831/5} = 0.2768$ . Treatment  $K$  reduced the number of wireworms significantly below all other treatments, but there is no indication of any difference between the other treatments. The treatment means may be reconverted to the original units by squaring.

**4. Experimental variation of the binomial type.** In this case the yields are obtained by examining a constant number  $n$  units per plot and noting those which possess a certain attribute (e.g., plants which are diseased). Experimental variation is presumed to arise solely from the binomial variation of the observed fraction  $p$  possessing the attribute about the expected fraction  $P$ , which is specified in terms of unknown parameters representing the treatment and soil effects.

If  $r_i$  is the number possessing the attribute on a typical plot, so that  $p_i = r_i/n$  the likelihood function takes the form

$$\prod_i \frac{n!}{r_i!(n-r_i)!} P_i^{r_i} Q_i^{n-r_i}.$$

Hence the terms in the logarithm which involve the unknown parameters are given by

$$(15) \quad L = \sum_i \{r_i \log P_i + (n - r_i) \log Q_i\}.$$

The equation of estimation for a typical constant  $\theta$  is

$$(16) \quad \sum \frac{n}{P_i Q_i} (p_i - P_i) \frac{\partial P_i}{\partial \theta} = 0$$

where the summation is over all plots whose expectations involve  $\theta$ .

As in the Poisson case, an exact solution is laborious because of the weights  $\frac{n}{P_i Q_i} \frac{\partial P_i}{\partial \theta}$ . The unequal weighting may be removed by transforming to the variate  $\alpha_i = \sin^{-1} \sqrt{P_i}$ , and assuming that the prediction formula is linear in the transformed scale. For a Latin square the prediction formula is assumed to be

$$(17) \quad \alpha_i = G + T_t + R_r + C_c$$

where the  $i$ th plot receives treatment  $t$  and lies in the  $r$ th row and  $c$ th column. Further

$$(18) \quad \sum T_t = \sum R_r = \sum C_c = 0.$$

Since  $P_i = \sin^2 \alpha_i$ ,  $\frac{dP_i}{d\alpha_i} = 2\sqrt{P_i Q_i}$ . A set of variates  $a'_i$  is defined so that on each plot

$$(19) \quad p_i - P_i = (a'_i - \alpha_i) \frac{dP_i}{d\alpha_i} = 2\sqrt{P_i Q_i} (a'_i - \alpha_i).$$

With these substitutions, the equation of estimation for  $T_i$ , for instance, becomes

$$(20) \quad \sum_{T_i} 4n(a'_i - \alpha_i) + \lambda = 0$$

where, as before,  $\lambda$  is an undetermined multiplier. The remainder of the solution proceeds exactly as in the Poisson case,  $T_i$  being found to be the difference between the mean value of  $a'_i$  over all plots receiving this treatment and the general mean of  $a'_i$ . A  $\chi^2$  test may be made with  $\sum_i 4n(a'_i - \alpha_i)^2$ .

From (19)

$$(21) \quad a'_i = \alpha_i + \frac{1}{2\sqrt{P_i Q_i}} (p_i - P_i) = \alpha_i + \frac{1}{2\sqrt{P_i Q_i}} (Q_i - q_i)$$

$$(22) \quad = \alpha_i + \frac{1}{2} \cot \alpha_i - q_i \operatorname{cosec} (2\alpha_i)$$

where  $q_i$  is the observed fraction which does not possess the attribute. The calculation of approximations to  $a'_i$  thus involves finding a predicted value  $\alpha_i$  from the treatment and block (or row and column) means, and using equation (22). Tables [10] of the values of  $\sin^{-1} \sqrt{P_i}$ ,  $\alpha_i + \frac{1}{2} \cot \alpha_i$ , and  $\operatorname{cosec} (2\alpha_i)$  have been prepared to facilitate the computations. It should be noted that these tables are in degrees, whereas the above equations assume that  $\alpha_i$  is measured in radians. In degrees, equation (20) above becomes

$$(23) \quad \sum_{T_i} \frac{\pi^2 n}{8100} (a'_i - \alpha_i) = 0$$

while

$$(24) \quad a'_i = \alpha_i + \frac{180}{\pi} \left\{ \frac{1}{2} \cot \alpha_i - q_i \operatorname{cosec} (2\alpha_i) \right\}.$$

As in the Poisson case, the appropriateness of the linearly additive law in equivalent angles depends on the way in which treatment and soil effects operate. As Bliss has shown [11], the effect of the transformation is to flatten out the cumulative normal frequency distribution, extending the range over which it can be approximated by a straight line.

**5. Numerical example of the angular transformation.** The data were selected from a randomized blocks experiment by Carruth [12] on the control by mechanical and insecticidal methods of damage due to corn ear worm larvae.

The control and the six types of mechanical protection were chosen for analysis, the "yields" being the percentages of ears unfit for sale. The numbers of ears varied somewhat from plot to plot, the average being 36.5, but the variations were fairly small and appeared to be random. It was considered that variations in the weight ( $4n$ ) could be ignored in solving the equations of estimation.

TABLE III  
*Percentages of unfit ears of corn*

Treatments	Blocks						Means
	I	II	III	IV	V	VI	
1	42.4 <sup>1</sup>	34.3	24.1	39.5	55.5	49.1	
	40.6 <sup>2</sup>	35.8	29.4	38.9	48.2	44.5	39.57 <sup>2</sup>
	40.7 <sup>3</sup>	36.0	29.4	38.9	48.6	44.6	39.70 <sup>3</sup>
2	23.5	15.1	11.8	9.4	31.7	15.9	
	29.0	22.9	20.1	17.9	34.3	23.5	24.62
	29.1	23.1	20.3	18.2	34.3	23.5	24.75
3	33.3	33.3	5.0	26.3	30.2	28.6	
	35.2	35.2	12.9	30.9	33.3	32.3	29.97
	35.5	35.3	14.5	31.0	33.4	32.4	30.35
4	11.4	13.5	2.5	16.6	39.4	11.1	
	19.7	21.6	9.1	24.0	38.9	19.5	22.13
	19.8	21.7	10.0	24.4	39.9	19.6	22.57
5	14.3	29.0	10.8	21.9	30.8	15.0	
	22.2	32.6	19.2	27.9	33.7	22.8	26.40
	22.6	32.7	19.2	28.0	33.7	22.9	26.52
6	8.5	21.9	6.2	16.0	13.5	15.4	
	17.0	27.9	14.4	23.6	21.6	23.1	21.27
	17.4	28.2	14.5	24.0	22.1	23.2	21.57
7	16.6	19.3	16.6	2.1	11.1	11.1	
	24.0	26.1	24.0	8.3	19.5	19.5	20.23
	24.3	26.2	28.8	10.9	20.1	19.5	21.63
Means	26.81 <sup>2</sup>	28.87	18.44	24.50	32.79	26.46	26.31

<sup>1</sup> Percentage. <sup>2</sup> Equivalent angle. <sup>3</sup> Second approximation.

The percentages of unfit ears, the equivalent angles and the second approximations to  $a'$  are shown in descending order in Table III. The percentages on

individual plots vary from 2.1 to 55.5. The second approximations were calculated from the block and treatment means of the angles. For the control plot (treatment 1) in block I, for example, the expected value is

$$39.57 + 26.81 - 26.31 = 40.07.$$

Since Fisher and Yates's tables of  $\alpha + \frac{1}{2} \cot \alpha$  and  $\operatorname{cosec} (2\alpha)$  are given for values of  $\alpha$  from  $45^\circ$  to  $90^\circ$ , we take the complement of the expected value, which is 49.93. Interpolating mentally from the table, we find

$$\alpha + \frac{1}{2} \cot \alpha = 74.0, \operatorname{cosec} (2\alpha) = 58.3.$$

Thus the second approximation to the complement of the angle is

$$74.0 - 0.424 \times 58.3 = 49.3.$$

Hence the second approximation to  $a'$  is 40.7, which agrees very closely with the equivalent angle.

On the majority of the plots, the second approximation differs by only a trivial amount from the equivalent angle. The plots with the three lowest percentages (2.1, 2.5, and 5.0) have increased somewhat more, and also one or two other plots where the angles deviated considerably from the expected values. A third set of approximations was not considered necessary.

The analysis of variance of the second approximations is given in Table IV.

TABLE IV

	Degrees of freedom	Sum of squares	Mean squares
Blocks	5	709.79	
Treatments	6	1,531.56	255.26
Error	30	982.67	32.76

Taking  $n$  as 36.5, the expected value of the error mean square is  $820.7/36.5 = 22.48$ . Thus  $\chi^2 = 982.67/22.48 = 43.71$ , with 30 degrees of freedom, which is almost exactly at the 5 percent level. This, together with the appreciable amount of the variance removed by blocks, indicates that the experimental error probably contains some element other than binomial variation. As in the preceding case, it would be wise to make the usual analysis of variance tests with the actual error mean square.

**6. Discussion.** It must be emphasized that the solutions given above apply to the case where the whole of the experimental error variation is of the Poisson or binomial type. The methods are therefore likely to be useful in practice only where the experimental conditions have been carefully controlled, or where the data are derived from such small numbers that the Poisson or binomial variation is much larger than any extraneous variation. The  $\chi^2$  test is helpful in deciding



whether this assumption is justified. Further, the examples worked above indicate that the transformed values form very good approximations on most plots. It will often be sufficient to adjust only those plots which give zero or very small values in the Poisson case, or zero or 100 percent values in the binomial case. In this connection the method of adjustment given above may perhaps be considered as an improvement on the empirical rule given by Bartlett [13] of counting  $n$  out of  $n$  as  $(n - 1/4)$  out of  $n$ .

Where extraneous variation becomes important, as is probably the normal case with data derived from field experiments, there seem to be no theoretical grounds for using the adjusted values. If we were prepared to describe accurately the nature of the variation other than that of the Poisson or binomial type, a new set of maximum likelihood equations could be developed. These would, however, lead to a different type of adjustment.

The justification for the use of transformations has no direct relation to the Poisson or binomial laws in this case, or in cases where percentages are derived from the ratios of two *weights* or volumes, as in chemical analyses, or from an arbitrary observational scoring. With percentages, for example, it may be said, without describing the experimental variation in detail, that the variance must vanish at zero and 100 percent and is likely to be greatest in the middle. The formula  $V = \lambda PQ$  is at least a first approximation to this situation. The angular transformation will approximately equalize a distribution of variances of this type, provided that  $\lambda$  is sufficiently small. We have, of course, returned to an "approximate" type of argument. It follows that the original data should be scrutinized carefully before deciding that a transformation is necessary and that any presumed opinions about the nature of the experimental variation should be verified as far as possible.

**7. Summary.** This paper discusses the theoretical basis for the use of the square root and inverse sine transformations in analyzing data whose experimental errors follow the Poisson and binomial frequency laws respectively.

The maximum likelihood equations of estimation are developed for each case, but are in general too complicated for frequent use. If, however, the expected yield of any plot is assumed to be an additive function of the treatment and soil effects in the transformed scale, a transformation can be found so that the equations of estimation assume the simple "normal theory" form. The transforms are closely related to the square roots and inverse sines respectively.

The nature of the assumed formula for the expected values is briefly discussed, and a  $\chi^2$  test is developed for the combined hypotheses that the prediction formula is satisfactory and that the experimental errors follow the assumed law.

Numerical examples are worked for both types of transformation. These indicate that even for data derived from small numbers, the square roots or inverse sines are good estimates of the correct transforms on almost all plots, except those which give zero yields in the Poisson case, or percentages near zero or 100 in the binomial case.

In practice, these new methods are not recommended to supplant the simple transformations for general use, because it can seldom be assumed that the whole of the experimental error variation follows the Poisson or binomial laws. The more exact analysis may, however, be useful (i) for cases in which the plot yields are very small integers or the ratios of very small integers (ii) in showing how to give proper weight to an occasional zero plot yield.

## REFERENCES

- [1] W. G. COCHRAN, "Some difficulties in the statistical analysis of replicated experiments," *Empire J. Expt. Agric.*, Vol. 6 (1938), pp. 157-75.
- [2] M. S. BARTLETT, "The square root transformation in the analysis of variance," *J. Roy. Stat. Soc. Suppl.*, Vol. 3 (1936), pp. 68-78.
- [3] C. I. BLISS, "The transformation of percentages for use in the analysis of variance," *Ohio J. Sci.*, Vol. 38 (1938), pp. 9-12.
- [4] A. CLARK AND W. H. LEONARD, "The analysis of variance with special reference to data expressed as percentages," *J. Amer. Soc. Agron.*, Vol. 31 (1939), pp. 55-56.
- [5] W. G. COCHRAN, "The  $\chi^2$  distribution for the Binomial and Poisson series, with small expectations," *Ann. Eugen.*, Vol. 7 (1936), pp. 207-17.
- [6] P. V. SUKHATME, "On the distribution of  $\chi^2$  in samples of the Poisson series," *J. Roy. Stat. Soc. Suppl.*, Vol. 5 (1938), pp. 75-9.
- [7] C. I. BLISS, "The determination of the dosage-mortality curve from small numbers," *Quart. J. Pharmacy and Pharmacology*, Vol. 11 (1938), pp. 192-216.
- [8] A. W. JONES, "Practical field methods of sampling soil for wireworms," *J. Agric. Res.*, Vol. 54 (1937), pp. 123-34.
- [9] W. G. COCHRAN, "The information supplied by the sampling results," *Ann. App. Biol.*, Vol. 25 (1938), pp. 383-9.
- [10] R. A. FISHER AND F. YATES, *Statistical tables for agricultural, biological and medical research*, Edinburgh, Oliver and Boyd, 1938.
- [11] C. I. BLISS, "The analysis of field experimental data expressed in percentages," *Plant Protection* (Leningrad), 1937, pp. 67-77.
- [12] L. A. CARRUTH, "Experiments for the control of larvae of *Heliothis Obsoleta* Fabr.," *J. Econ. Ent.*, Vol. 29 (1936), pp. 205-9.
- [13] M. S. BARTLETT, "Some examples of statistical methods of research in agriculture and applied biology," *J. Roy. Stat. Soc. Suppl.*, Vol. 4 (1937), p. 168, footnote.

IOWA STATE COLLEGE,  
AMES, IOWA

## NOTES

*This section is devoted to brief research and expository articles, notes on methodology and other short items.*

### ORTHOGONAL POLYNOMIALS APPLIED TO LEAST SQUARE FITTING OF WEIGHTED OBSERVATIONS

BY BRADFORD F. KIMBALL

**1. Introduction.** Let the independent variable be denoted by  $x$ , and let it range over  $n$  consecutive integral values  $x_1$  to  $x_n$ . Thus  $x$  represents the index-number of the ordered intervals at which observations are taken, where the intervals are all of equal length, and an index-number is assigned in consecutive order to every interval within the range of investigation, *whether observations occur in that interval or not*. Let  $y_x$  denote the observation measure (usually referred to as observed value), if such observation exists. Let  $w_x$  denote the weight of that observation, with weight zero assigned where observations are lacking.

To shorten the notation, summation over all values of  $x$  from  $x_1$  to  $x_n$  will be denoted by the sign  $\Sigma$ . If a subscript and superscript is used, the context will indicate the variable to which the summation refers. The  $r$ th binomial coefficient will be denoted by  $\binom{x}{r}$ .

A system of polynomials  $\phi_r(x)$ ,  $r = 0, 1, 2, 3, \dots$  of degree  $r$  in  $x$  is said to be an orthogonal system, for the purposes of this paper, if they satisfy the relations

$$(1) \quad \sum W_x \phi_r(x) \phi_s(x) \quad \begin{cases} = 0, & r \neq s \\ \neq 0, & r = s. \end{cases}$$

To construct the polynomials, one may write them in the form

$$(2) \quad \begin{aligned} \phi_0(x) &= f_0(x) = \text{constant} \\ \phi_r(x) &= f_r(x) - \sum_{i=0}^{r-1} h_i \phi_i(x) \quad r = 1, 2, 3, \dots \end{aligned}$$

where the  $h_i$  are constants and the  $f_r(x)$  are arbitrary polynomials of degree  $r$ . It then follows from the conditions of orthogonality that

$$(3) \quad h_i = \frac{\sum w_x f_r(x) \phi_i(x)}{\sum w_x [\phi_i(x)]^2}.$$

Thus when the polynomials  $f_r(x)$  have been chosen for all  $r$ , the system of orthogonal polynomials for a given set of weights can be constructed and is uniquely determined except for a constant factor [1].

By virtue of the relation (2) and the conditions of orthogonality (1), it follows that

$$(4) \quad \Sigma w_x [\phi_r(x)]^2 = \Sigma w_x f_r(x) \phi_r(x).$$

Define the function  $\Phi(r, k)$  by

$$(5) \quad \Phi(r, k) = \Sigma w_x f_r(x) \phi_k(x), \quad r = 0, 1, 2, 3, \dots$$

It follows from the relations (2) and (3) that

$$(6) \quad \phi_r(x) = f_r(x) - \sum_{i=0}^{r-1} \frac{\Phi(r, i)}{\Phi(i, i)} \phi_i(x)$$

where it is to be noted that this summation is independent of  $x$ .

Define  $q_r$  and  $Y_r$  by

$$(7) \quad q_r = \Sigma w_x [\phi_r(x)]^2 = \Sigma w_x f_r(x) \phi_r(x) = \Phi(r, r),$$

$$(8) \quad Y_r = \Sigma w_x y_x \phi_r(x).$$

Then if  $u_r(x)$  represents the polynomial solution of degree  $r$  of the normal equations set up for observed values  $y_x$  and weights  $w_x$ ,

$$(9) \quad u_r(x) = \frac{Y_0}{q_0} + \frac{Y_1}{q_1} \phi_1(x) + \frac{Y_2}{q_2} \phi_2(x) + \dots + \frac{Y_r}{q_r} \phi_r(x).$$

If  $E^2$  denotes the weighted sum of the squares of the discrepancies between the ordinates  $u_r(x)$  of the fitted curve and the observed values  $y_x$ , then [2],

$$(10) \quad E^2 = \Sigma w_x [u_r(x) - y_x]^2 = \Sigma w_x y_x^2 - \sum_{i=0}^r \frac{Y_i^2}{q_i}.$$

The practicability of the use of orthogonal polynomials is thus seen to depend upon whether the quantities  $\Phi(r, k)$  and  $Y_r$  can be evaluated in a reasonably simple manner.

The thesis of this paper is that if  $f_r(x)$  is taken as the binomial coefficient  $\binom{x}{r}$ , one can effectively apply the method of orthogonal polynomials. This is made possible by the use of factorial moments in conjunction with an adding machine that prints cumulative totals.

In treating the same problem Aitken sets up the normal equations in terms of factorials, but considers the explicit use of orthogonal polynomials impractical. He writes: "the arbitrary nature of the weights stands in the way of any analytical sophistication; orthogonal polynomials emerge, but are not of great use; and the necessity of solving the moment equations cannot be circumvented" [3]. He prefers a determinantal method of solution of the normal

equations which the writer has found to be more involved from a practical point of view, than the present method, although it is elegant from a theoretical standpoint.

Thus although the present method is not new from the point of view of theory, the writer has found that forms made up by the use of the technique suggested below, offer an effective method for fitting polynomial curves to weighted observations.

**2. Simplification of the problem when  $f_r(x) = \binom{x}{r}$ .** Factorial moments  $S_r$  and  $M_r$  are defined by

$$(11) \quad S_r = \sum \binom{x}{r} w_x, \quad M_r = \sum \binom{x}{r} w_x y_x \quad r = 0, 1, 2, \dots$$

These moments are not difficult to compute and are readily checked as computed. Formula for  $\Phi(r, k)$  then becomes

$$(12) \quad \Phi(r, k) = \sum \binom{x}{r} w_x \phi_k(x).$$

Thus since  $\phi_0(x) = 1$ ,  $\Phi(r, 0) = \sum \binom{x}{r} w_x = S_r$  and hence

$$\phi_1(x) = \binom{x}{1} - \frac{\Phi(1, 0)}{\Phi(0, 0)} = x - \frac{S_1}{S_0}.$$

Again

$$\begin{aligned} \Phi(r, 1) &= \sum \binom{x}{r} w_x \left( x - \frac{S_1}{S_0} \right) = \sum \binom{x}{r} \binom{x}{1} w_x - \frac{S_1}{S_0} \sum \binom{x}{r} w_x \\ &= (r+1)S_{r+1} + rS_r - \frac{S_1 S_r}{S_0}. \end{aligned}$$

Hence

$$q_1 = \Phi(1, 1) = 2S_2 + \left( 1 - \frac{S_1}{S_0} \right) S_1.$$

A recursion formula for  $\Phi(r, k)$  may be obtained by expanding  $\phi_k(x)$  in formula (12) by means of (6). Thus

$$\begin{aligned} (13) \quad \Phi(r, k) &= \sum \binom{x}{r} \binom{x}{k} w_x - \sum_{i=0}^{k-1} \frac{\Phi(k, i)}{q_i} \left[ \sum \binom{x}{r} w_x \phi_i(x) \right] \\ &= \sum \binom{x}{r} \binom{x}{k} w_x - \sum_{i=0}^{k-1} \frac{\Phi(r, i) \Phi(k, i)}{q_i}. \end{aligned}$$

The first term can be easily expressed as a linear combination of binomial coefficients, and thus as a linear combination of moments  $S_i$ .

The formula for  $Y_r$  can be broken down as follows:

$$\begin{aligned}
 Y_0 &= \sum w_s y_s = M_0, \\
 (14) \quad Y_r &= \sum w_s y_s \phi_r(x) = \sum w_s y_s \binom{x}{r} - \sum_{i=0}^{r-1} \frac{\Phi(r, i)}{q_i} [\sum w_s y_s \phi_i(x)] \\
 M_r &- \sum_{i=0}^{r-1} \frac{\Phi(r, i)}{q_i} Y_i.
 \end{aligned}$$

Thus

$$\begin{aligned}
 Y_1 &= M_1 - \frac{S_1}{S_0} Y_0, \\
 Y_2 &= M_2 - \frac{\Phi(2, 1)}{q_1} Y_1 - \frac{\Phi(2, 0)}{q_0} Y_0, \quad \text{etc.}
 \end{aligned}$$

**3. General technique of computation.** In determining the best fitting polynomial of degree  $r$ , the ratios  $\Phi(r, i)/q_i$  are seen to play an important part. In a form for calculation, these quantities should receive simple designations such as  $b_i$  for a second degree curve,  $c_i$  for a third degree curve, etc. Suppose they are designated by  $R_i$  for a curve of degree  $r$ ; then

$$(15) \quad \phi_r(x) = \binom{x}{r} - \sum_{i=0}^{r-1} R_i \phi_i(x)$$

$$(16) \quad Y_r = M_r - \sum_{i=0}^{r-1} R_i Y_i$$

$$(17) \quad q_r = \sum \binom{x}{r}^2 w_s - \sum_{i=0}^{r-1} R_i \Phi(r, i)$$

and in determining  $\Phi(r, k)$  for  $k = 0, 1, 2, \dots, r-1$ , formula (13) may be written:

$$(18) \quad \Phi(r, k) = \sum \binom{x}{r} \binom{x}{k} w_s - \sum_{i=0}^{k-1} R_i \Phi(k, i).$$

The fact that these quantities  $R_i$  appear as multipliers in so many of the fundamental formulas greatly simplifies the mechanics of the calculation, especially when a calculating machine is used.

In final determination of polynomial curve the differences of the polynomial at  $x = 0$  are readily determined since the leading term of each orthogonal polynomial is a binomial coefficient and thus

$$\begin{aligned}
 (19) \quad \Delta^k \phi_r(0) &= - \sum_{i=0}^{r-1} R_i \Delta^k \phi_i(0), \quad k = 1, 2, 3, \dots, r-1 \\
 \Delta^r \phi_r(0) &= 1.
 \end{aligned}$$

Since the effectiveness of the method depends upon the availability of an adding machine which records a cumulative subtotal, the determination of the curve from the differences at the point  $x = 0$  is not a hardship and indeed affords a quick and accurate means of setting up the curve for purposes of plotting and checking.

$$u_r(0) = \frac{Y_0}{q_0} + \frac{Y_1}{q_1} \phi_1(0) + \frac{Y_2}{q_2} \phi_2(0) + \dots + \frac{Y_r}{q_r} \phi_r(0),$$

$$(20) \quad \Delta^k u_r(0) = \frac{Y_k}{q_k} + \frac{Y_{k+1}}{q_{k+1}} \Delta^k \phi_{k+1} + \dots + \frac{Y_r}{q_r} \Delta^k \phi_r(0),$$

$$\Delta^r u_r(0) = \frac{Y_r}{q_r}.$$

The advantage of the use of orthogonal polynomials becomes particularly apparent when error formulae are to be used. The formula for the sum of the squares of the discrepancies, denoted by  $E^2$ , is given above (formula (10)). The estimated variance  $V$  of the weighted observations about the fitted curve is thus  $E^2/(n - r - 1)$  where  $n$  is the number of values of  $x$  used in fitting and  $r$  is the degree of the curve fitted. Recalling that the matrix of the normal equations is of the diagonal form with diagonal elements  $q_0, q_1, \dots, q_r$  it follows that the coefficient  $Y_k/q_k$  of  $\phi_k(x)$  in the expansion of  $u_r(x)$  has the variance  $V/q_k$ .

Furthermore the variance of the ordinate of the fitted curve  $u_r(x)$  at a point  $x$  due to sampling variations in the determination of the coefficients of the curve, under the assumption that the weights and values of the independent variable  $x$  do not involve errors, has the simple form

$$(21) \quad \begin{array}{l} \text{Variance of } u_r(x) \\ \text{at point } x = \end{array} V \left[ \frac{\phi_0^2(x)}{q_0} + \frac{\phi_1^2(x)}{q_1} + \dots + \frac{\phi_r^2(x)}{q_r} \right]$$

since the covariances of the orthogonal polynomials are zero [4].

#### REFERENCES

- [1] G. SZEGO, *Orthogonal Polynomials*, Colloquium Publications, Amer. Math. Soc., New York, 1939, Vol. 23, Chapter II.
- [2] MAX SASULY, *Trend Analysis of Statistics*, The Brookings Institution, Washington, 1934, pp. 296-297.
- [3] A. C. AITKEN, "On Fitting Polynomials to Weighted Data by Least Squares," *Proc. Roy. Soc. of Edinburgh*, Session 1933-34, Vol. 54, Part I, No. 1, p. 2.
- [4] HENRY SCHULTZ, "The Standard Error of a Forecast from a Curve," *Jour. Amer. Stat. Assn.*, June 1930; or Whittaker & Robinson, *The Calculus of Observations*, Blackie & Son Ltd., London, 1937, 2d ed., p. 242.

# COMBINATORIAL FORMULAS FOR THE $r$ th STANDARD MOMENT OF THE SAMPLE SUM, OF THE SAMPLE MEAN, AND OF THE NORMAL CURVE

By P. S. DWYER

The standard moments of the normal curve are usually expressed by the two statements [1, p. 97]

$$(1) \quad \begin{cases} \alpha_{2s} &= \frac{(2s)!}{2^s s!} \\ \alpha_{2s+1} &= 0 \end{cases}$$

It is of some interest to note that these two statements may be generalized into a single statement by observing that  $\frac{(2s)!}{2^s s!}$  is the number of ways in which  $2s$  things can be grouped in pairs and that 0 is the number of ways in which  $2s + 1$  things can be grouped in pairs. It is obvious that an odd number of things can not be grouped in pairs since there must be at least one unpaired unit. It is clear, too, that the number of orders in which  $2s$  things can be grouped in pairs is  $\binom{2s}{2} \binom{2s-2}{2} \binom{2s-4}{2} \cdots \binom{4}{2} \binom{2}{2}$  and this is  $\frac{(2s)!}{2^s}$ . However if the resulting paired groups (rather than the orders of grouping) are counted it is seen that each paired grouping is repeated  $s!$  times so that  $\frac{(2s)!}{2^s s!}$  represents the number of ways  $2s$  things can be grouped in pairs. If we arbitrarily define the number of ways 0 things can be grouped in pairs to be 1 (or if we limit our theorem to values of  $r > 0$ ) we may say "The  $r$ th standard moment of the normal curve is equal to the number of ways in which  $r$  things can be grouped in pairs."

As presented above the combination representation is used primarily as a means of unification of results. However, it is possible to derive the standard moments of the normal curve in such a way as to indicate the term  $\frac{(2s)!}{2^s s!}$  early in the proof and to trace it throughout the proof. I follow the method outlined by H. C. Carver [2] in obtaining the normal distribution as the limit of the distribution of sample sums (or of sample means) though I use a somewhat different notation [3, p. 5]. If we let  $\binom{1^r}{p_1^{r_1} \cdots p_s^{r_s}}$  represent the number of ways in which  $r$  units can be collected with  $\pi_1$  groups containing  $p_1$  units,  $\pi_2$  groups containing  $p_2$  units, etc., then the multinomial theorem can be expressed as [3, p. 17]

$$(2) \quad (1)^r = \sum \binom{1^r}{p_1^{r_1} \cdots p_s^{r_s}} (p_1^{r_1} \cdots p_s^{r_s})$$



where the summation is taken over all possible partitions  $p_1^{r_1} \dots p_s^{r_s}$  of  $r$  and the expression  $(p_1^{r_1} \dots p_s^{r_s})$  represents the power product form [3, p. 14] which is  $\pi_1! \pi_2! \dots \pi_s!$  times the monomial symmetric function. If  $\rho$  represents the number of parts of the partition then

$$\rho = \pi_1 + \pi_2 + \dots + \pi_s$$

while

$$r = p_1 \pi_1 + p_2 \pi_2 + \dots + p_s \pi_s.$$

Now it can be shown from (2) in the case of infinite sampling that

$$(3) \quad \bar{\mu}_{r:(1)} = \sum \binom{1^r}{p_1^{r_1} \dots p_s^{r_s}} n^{(\rho)} (\bar{\mu}_{p_1})^{\pi_1} \dots (\bar{\mu}_{p_s})^{\pi_s}$$

and since  $\bar{\mu}_1 = 0$ , it is only necessary to sum over all partitions which have no unit part. We have then, dividing by  $[\bar{\mu}_{2:(1)}]^{1/2} = [n\bar{\mu}_2]^{1/2}$

$$(4) \quad \alpha_{r:(1)} = \sum \binom{1^r}{p_1^{r_1} \dots p_s^{r_s}} \frac{n^{(\rho)}}{n^{1/2}} (\alpha_{p_1})^{\pi_1} \dots (\alpha_{p_s})^{\pi_s}.$$

We have now a formula for the  $r$ th standard moment of the sample sum which is expressed essentially in combination notation since the quantity  $\binom{1^r}{p_1^{r_1} \dots p_s^{r_s}}$  represents the number of ways in which  $r$  units can be grouped to form  $\pi_1$  groups containing  $p_1$  units,  $\pi_2$  groups containing  $p_2$  units, etc. All non-unitary groupings of  $r$  are formed, each combinatorial coefficient is computed and multiplied by  $n^{(\rho)}/n^{1/2}$  times the product of the corresponding  $\alpha$ 's, and the sums are formed. It might be noted that the formula for the  $r$ th standard moment of the sample mean is identical with (4) while the corresponding finite sampling (without replacements) formula is

$$(5) \quad \alpha_{r:(1)} = \sum \binom{1^r}{p_1^{r_1} \dots p_s^{r_s}} \frac{N^\rho P_{p_1^{r_1} \dots p_s^{r_s}}}{N^{1/2} P_1^{1/2}} (\alpha_{p_1})^{\pi_1} \dots (\alpha_{p_s})^{\pi_s}.$$

The  $P$ 's are defined in previous papers [2, p. 105-6][3, p. 113].

We obtain the formula for the  $r$ th standard moment of the normal curve by taking the limit of (4) as  $n \rightarrow \infty$ . (H. C. Carver has pointed out [2, p. 121] that this method of derivation imposes fewer restrictions than does the derivation from Hagen's hypothesis.) Each partition term will approach zero as  $n$  approaches infinity if  $\rho < \frac{1}{2}r$ . Now the only non-unitary partition in which  $\rho$  is not less than  $\frac{1}{2}r$  is the partition  $2^{1/2}r$  and we can have this partition only when  $r$  is even. Now the limit as  $n$  approaches infinity of  $n^{(\rho)}/n^{1/2}$  is unity and we have, in the limiting case

$$(6) \quad \alpha_r = \begin{cases} \binom{1^r}{2^{1/2}r} & \text{if } r \text{ is even.} \\ 0 & \text{if } r \text{ is odd.} \end{cases}$$

Since  $\binom{1^r}{2^r}$  is the number of ways  $r$  units can be grouped in pairs when  $r$  is even and since 0 is the number of ways  $r$  units can be grouped in pairs where  $r$  is odd, it follows that the  $r$ th standard moment of the normal curve is the number of ways in which  $r$  units can be grouped in pairs.

This development is of interest in that it makes possible the tracing of the value  $\binom{1^r}{2^r}$  back through the various stages of the development to the coefficient of  $(2^r)$  in the power product expansion of the multinomial theorem.

#### REFERENCES

- [1] H. C. CARVER, "Frequency Curves," Chapter VII of *Handbook of Mathematical Statistics*.
- [2] H. C. CARVER, "Fundamentals of the theory of sampling," *Annals of Math. Stat.*, Vol. 1 (1930), pp. 101-121.
- [3] P. S. DWYER, "Combined Expansions of Products of Symmetric Power Sums and of Sums of Symmetric Power Products with Application to Sampling," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 1-47, 97-132.

UNIVERSITY OF MICHIGAN,  
ANN ARBOR, MICHIGAN

### ON A METHOD OF SAMPLING<sup>1</sup>

BY E. G. OLDS

It is recorded that Diogenes fared forth with a lantern in his search for an honest man. History does not tell us how many dishonest men he encountered before he found the first honest one but, judging from the fact that he took his lantern, apparently he expected to have a long search. The general problem of sampling inspection, of which the above is a special case, can be stated as follows:

Given a lot, of size  $m$ , containing  $s$  items of a specified kind. If items are to be drawn without replacement until  $i$  of the  $s$  items have been drawn, how many drawings, on the average, will be necessary?

Uspensky<sup>2</sup> has solved a problem concerning balls in an urn, from which the answer to the above question can be obtained for the special case  $i = 1$ . For the general case, the distribution for the number  $n$  of the drawing in which the  $i$ th specified item appears, is given by terms of the series:

$$(1) \quad \nu'_0 = \sum_{n=0}^{m-s+i} \frac{C_{n-1,i-1} C_{m-n,s-i}}{C_{m,s}} = \sum_{n=0}^{\infty} C_{n-1,i-1} C_{m-n,s-i}$$

<sup>1</sup> Presented to The Institute of Mathematical Statistics, Dec. 27, 1938, at Detroit, Mich., as part of a paper, entitled "Remarks on two methods of sampling inspection."

<sup>2</sup> J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937, p. 178.

where the first symbol indicates the number of ways of choosing  $i - 1$  of the specified items to fill the first  $n - 1$  places, the second symbol indicates the number of ways of disposing of  $s - 1$  specified items in the last  $m - n$  places, and the denominator gives the number of ways that the  $s$  items can be scattered through the lot. In order to get the average number of draws we multiply  $\nu'_0$  by  $n$  and sum. Then we have

$$(2) \quad \nu'_1 = \sum_{n=0}^{\infty} \frac{n C_{n-1, i-1} C_{m-n, s-1}}{C_{m, s}} = \frac{i(m+1)}{s+1} \sum_{n=0}^{\infty} \frac{C_{n, i} C_{m-n, s-1}}{C_{m+1, s+1}} = \frac{i(m+1)}{s+1}.$$

*Example 1.* On a table of 200 bargain shirts there are 5 which have a 15 in. neckband and 35 in. sleeves. How many shirts must be examined, on the average, to find two of the desired kind?

*Solution.* For this case,  $m = 100$ ,  $s = 5$ ,  $i = 2$ . Therefore  $\bar{n} = [2(201)] \div 6 = 67$ . Thus, an average of 67 shirts must be examined.

Suppose  $\mu_K$  represents the  $K$ th moment about the mean,  $\nu_K$  the  $K$ th moment about the origin, and  $\nu'_K$  the moment relation given by

$$(3) \quad \nu'_K = (\nu_1 + K - 1)^{(K)},$$

where  $(\nu_1 + K - 1)^{(K)}$  represents the result of expanding  $(\nu + K - 1)^{(K)}$  and changing the exponent of  $\nu$  to the corresponding subscript. (For example,  $\nu'_3 = (\nu_1 + 2)^{(3)} = \nu_3 + 3\nu_2 + 2\nu_1$ .) It is easy to derive the recurrence relation

$$(4) \quad \nu'_K = \frac{(i + K - 1)(m + K)}{s + K} \nu'_{K-1}.$$

From this result the computation of the moments about the mean is theoretically direct. Actually the results do not seem to be very compact. The variance is given by

$$(5) \quad \mu_2 = \frac{(m+1)(m-s)}{(s+1)^2(s+2)} [i(s+1) - i^2].$$

In case  $s$  is unknown and  $n$  is known for a particular value of  $i$ , we may estimate  $s$ , (or rather  $\frac{1}{s+1}$ ), by using the relation,  $n = \frac{i(m+1)}{s+1}$ . Then

$$(6) \quad \frac{1}{s+1} \text{ est.} = \frac{n}{i(m+1)},$$

and the variance, using this estimate, is given by

$$(7) \quad \text{Variance of } \left( \frac{1}{s+1} \right) \text{ est.} = \frac{n}{n + i(m+1)} \cdot \frac{1}{i(m+1)} \left[ \frac{n}{i} - 1 \right] \left[ 1 - \frac{n}{m+1} \right].$$

*Example 2.* In order to check a box of 144 screws, screws are drawn until 10 good screws are obtained. In a particular case only 10 drawings were necessary. Estimate the number of good screws in the lot.

*Solution.* Here  $m = 144$ ,  $i = 10$ ,  $n = 10$ . The estimate for  $s$  is obtained

from  $\left(\frac{1}{s+1}\right)$  est. =  $\frac{10}{10(145)} = \frac{1}{145}$  and, as might be expected, the conclusion is that all the screws are good. Furthermore the variance of the estimated quantity is zero.

It is obvious that the number of draws necessary to obtain any particular number of specified items is correlated with the numbers of draws for lesser numbers of items. To investigate this, let us suppose that  $n_j$  represents the number of draws to obtain exactly  $j$  specified items and that  $x_j = n_j - n_{j-1}$ . It follows immediately from our previous results, that

$$(8) \quad E(x_1) = E(x_2) = E(x_3) = \dots = \frac{m+1}{s+1}.$$

This result could be obtained from the fact that, corresponding to any arrangement of the lot for which  $x_a = a$  and  $x_b = b$ , there is another arrangement where  $x_a = b$  and  $x_b = a$ , formed by moving  $a - b$  of the non-specified items from the first group to the second. From this fact we see, also, that

$$(9) \quad E(x_1^2) = E(x_2^2) = E(x_3^2) = \dots$$

$$\text{But } x_1 = n_1 \text{ and } \sigma_{n_1}^2 = \frac{(m+1)(m-s)}{(s+1)^2(s+2)} [s+1-1] = ds.$$

Therefore,

$$(10) \quad \sigma_{x_1}^2 = \sigma_{x_2}^2 = \dots = ds.$$

But, from our previous formula we have

$$\sigma_{n_2}^2 = d(2s-2), \quad \sigma_{n_3}^2 = d(3s-6), \text{ etc.}$$

Since  $n_2 = x_1 + x_2$ , it follows that

$$\sigma_{n_2}^2 = \sigma_{x_1}^2 + 2r_{x_1, x_2}\sigma_{x_1}\sigma_{x_2} + \sigma_{x_2}^2$$

where  $r_{x_1, x_2}$  is the correlation between  $x_1$  and  $x_2$ . Therefore,

$$(11) \quad r_{x_1, x_2} = -1/s.$$

Also, since  $x_1 = n_2 - x_2$ , it follows that

$$(12) \quad r_{n_2, x_2} = \sqrt{\frac{s-1}{2s}}.$$

Likewise, from  $x_2 = n_3 - x_1$ , we get

$$(13) \quad r_{n_3, x_1} = \sqrt{\frac{s-1}{2s}}.$$

Finally, we obtain the three general results

$$(14) \quad r_{n_i, x_{i+1}} = -\sqrt{\frac{i}{s(s-i+1)}},$$

$$(15) \quad r_{n_i, x_i} = \sqrt{\frac{s-i+1}{st}},$$

$$(16) \quad r_{n_{i+1}, n_i} = \sqrt{\frac{i(s-i)}{(i+1)(s-i+1)}}.$$

*Example 3.* The cards of a deck are turned one by one until two aces have appeared. The second ace appears when the 36th card is turned. How many more cards should one expect to have to turn to find a third ace?

*Solution.* Here  $m = 52$ ,  $s = 4$ ,  $i = 2$ ,  $n_2 = 36$ .

Then  $\bar{n}_2 = 2 \cdot \frac{53}{5}$ ,  $\bar{x}_3 = \frac{53}{5}$ , and  $r_{n_2, x_3} = -\sqrt{\frac{2}{4(4-2+1)}} = -\frac{\sqrt{6}}{6}$ . Also

$\sigma_{x_3} = \sqrt{4d}$  and  $\sigma_{n_2} = \sqrt{6d}$ . Since  $\frac{x_3 - \bar{x}_3}{\sigma_{x_3}} = r_{n_2, x_3} \frac{(n_2 - \bar{n}_2)}{\sigma_{n_2}}$ , we have

$$x_3 = \frac{53}{5} - \frac{2}{\sqrt{6}} \cdot \frac{\sqrt{6}}{6} \left( 36 - \frac{106}{5} \right) = \frac{17}{3}.$$

Of course this result could have been obtained more directly by noting that there were two aces left among the 16 remaining cards.

**Conclusion.** The results given in this note might be useful when it is necessary to estimate the number of items to be drawn in order to secure a desired number of a particular type, such as may be the case in obtaining a sample with previously defined characteristics. Also the note disproves such intuitive notions as the one that when looking for a desired record, one is most likely to have to search the whole pile to find it. As far as methods of sampling inspection are concerned, the one implied in this note has little to recommend it.

CARNEGIE INSTITUTE OF TECHNOLOGY,  
PITTSBURGH, PA.

## RANK CORRELATION WHEN THERE ARE EQUAL VARIATES<sup>1</sup>

BY MAX A. WOODBURY

If there is given a set of number pairs

$$(1) \quad (X_1, Y_1), (X_2, Y_2), \dots, (X_N, Y_N),$$

we may assign to each variate its "rank" (i.e. one more than the number of corresponding variates in the set greater than the given variate). In this way there is obtained a set of pairs of ranks

$$(2) \quad (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N).$$

<sup>1</sup> Presented at the fall meeting, Mich. section of the Math. Assn. of America, Nov. 18, 1939, Kalamazoo College.

If we assume that  $X_i \neq X_j$  and  $Y_i \neq Y_j$  when  $i \neq j$  then it follows that each integer from 1 to  $N$  appears once and only once in the  $x$ 's and the same holds for the  $y$ 's. This leads at once to the formulas:

$$(3a) \quad \sum_{i=1}^N x_i = \sum_{i=1}^N y_i = \sum_{i=1}^N i = N(N+1)/2,$$

$$(3b) \quad \sum_{i=1}^N x_i^2 = \sum_{i=1}^N y_i^2 = \sum_{i=1}^N i^2 = N(N+1)(2N+1)/6.$$

When these results are substituted in the expression for the product moment correlation coefficient we have after simplifying [1],

$$(4) \quad \rho = 1 - 6 \sum_{i=1}^N D_i^2 / N(N^2 - 1) \quad \text{where } D_i = x_i - y_i.$$

If we consider the case of equal variates and follow the rule for assigning ranks given in the first paragraph, the resulting method is known as the bracket-rank method. The use of (4) in the calculation of  $\rho$  by this method is not strictly valid, because not every integer appears in the summations and so neither (3a) nor (3b) is true.

The more accurate mid-rank method assigns to each of the equal variates the average of the ranks that would be assigned if we were to give them an arbitrary order. This method preserves (3a) but not (3b). In this paper  $\rho_M$  indicates the value of  $\rho$  as calculated by (4) when the mid-rank method is used.

In a method due to DuBois [2], the equal variates are assigned the same rank so as to satisfy (3b). In this case (3a) is not satisfied.

If we assign the ranks to the equal variates in an arbitrary way, then (3a) and (3b) are of course satisfied and the use of (4) is valid. There are two disadvantages to such a method; first, the equal variates are treated differently, and second, the assignment of ranks is arbitrary. These difficulties are removed if one uses the average of the values of  $\rho$  corresponding to all possible ways of arbitrarily assigning ranks to the equal variates. Since  $\rho$  is linear in  $\sum_i D_i^2$  the average value of  $\rho$  may be obtained from the average value of  $\sum_i D_i^2$  and the use of (4).

Let us first consider the simple case of two equal variates in one of the variables, say  $X$ . It is clear that there are only two possible ways of assigning ranks, and that if we arrange the series by the assigned  $x$  ranks, the resulting series differ only in the  $y$  ranks corresponding to the equal  $X$  variates. If we denote the two  $x$  ranks to be assigned by  $m$  and  $m+1$  and the  $y$ 's corresponding for a particular arrangement by  $y_m$  and  $y_{m+1}$  we have for the average  $\sum_i D_i^2$  the expression

$$(5a) \quad \sum_{i=1}^{m-1} (x - y_i)^2 + \sum_{i=m+2}^N (x - y_i)^2 \\ + \frac{1}{2}[(m - y_m)^2 + (m + 1 - y_{m+1})^2 + (m - y_{m+1})^2 + (m + 1 - y_m)^2].$$

By the mid-rank method the corresponding expression is

$$(5b) \quad \sum_{z=1}^{m-1} (x - y_z)^2 + \sum_{z=m+2}^N (x - y_z)^2 + (m + \frac{1}{2} - y_m)^2 + (m + \frac{1}{2} - y_{m+1})^2.$$

The correction  $\Delta_2$  to be added to the mid-rank  $\sum_i D_i^2$  to get the average  $\sum_i D_i^2$  is, by subtracting (5b) from (5a) and simplifying,

$$(6) \quad \Delta_2 = \frac{1}{2}.$$

To get  $\Delta_K$  in the more general case of several equal variates, we need only consider the difference between the average value of  $\sum_i D_i^2$  and that obtained by the mid-rank method. If there are  $K$  equal  $X$  variates we may assign the ranks in  $K!$  ways, this results in  $K!$  permutations of the  $y$  ranks for the sets arranged in order of their assigned  $x$  ranks. In  $(K-1)!$  permutations  $y_{m+i}$  corresponds to the  $x$  rank of  $m+i$  so that the correction to the mid-rank  $\sum_i D_i^2$  is

$$(7) \quad \begin{aligned} \Delta_K &= \frac{(K-1)!}{K!} \left[ \sum_{j=0}^{K-1} \sum_{i=0}^{K-1} (m+i-y_{m+i})^2 \right] - \sum_{j=0}^{K-1} \left( m + \frac{K-1}{2} - y_{m+j} \right)^2 \\ &= \frac{1}{K} \sum_{j=0}^{K-1} \sum_{i=0}^{K-1} \left[ (m+i-y_{m+i})^2 - \left( m + \frac{K-1}{2} - y_{m+j} \right)^2 \right] = \frac{K(K^2-1)}{12}. \end{aligned}$$

It is to be noticed that the correction is positive and depends *only* on the number of equal  $X$  variates. From this it can be concluded that for more than one group of equal variates no matter whether  $X$ 's or  $Y$ 's we can obtain the average  $\sum_i D_i^2$  by computing a correction for each group and then adding these corrections to get the total correction to the mid-rank  $\sum_i D_i^2$ . Then as before noted we can by (4) calculate the average  $\rho$  (denoted as  $\bar{\rho}$ ).

This correction to  $\sum_i D_i^2$  may be converted into a correction to  $\rho_M$ . That is

$$\text{if } \delta_{N,K_i} = \frac{6\Delta_{K_i}}{N(N^2-1)} = \frac{K_i(K_i^2-1)}{2N(N^2-1)}, \text{ then}$$

$$(8) \quad \bar{\rho} = \rho_M - \sum_i \delta_{N,K_i},$$

where the summation extends over all groups of equal variates, and  $K_i$  is the number of equal variates in the  $i$ th group.

A table of  $\delta_{NK}$  for different values of  $N$  and  $K$  is given, and also a table of  $\Delta_K$ . The values  $\Delta_K$  are given in the top row of the table, while the  $\delta_{NK}$  are given in the rows below.

Table of  $\Delta_K$  and  $\delta_{NK}$ 

$\begin{matrix} K \\ \backslash \\ N \end{matrix}$	2	3	4	5	6	7	8	9	10	11	12	13
$\Delta_K$	0.5000	2.000	5	10	17.5	28	42	60	82.5	110	143	182
$\delta_{NK}$												
3	1250	—	—	—	—	—	—	—	—	—	—	—
4	0500	2000	—	—	—	—	—	—	—	—	—	—
5	0250	1000	2500	—	—	—	—	—	—	—	—	—
6	0143	0571	1429	2857	—	—	—	—	—	—	—	—
7	0089	0357	0893	1786	3125	—	—	—	—	—	—	—
8	0060	0238	0595	1190	2083	3333	—	—	—	—	—	—
9	0042	0166	0417	0833	1458	2333	3500	—	—	—	—	—
10	0030	0121	0303	0606	1061	1697	2546	3636	—	—	—	—
11	0023	0091	0227	0455	0795	1273	1909	2727	3750	—	—	—
12	0017	0070	0175	0350	0612	0979	1469	2098	2885	3846	—	—
13	0014	0055	0137	0275	0480	0769	1154	1648	2266	3022	3929	—
14	0011	0044	0110	0220	0385	0615	0923	1319	1813	2418	3143	4000
15	0009	0036	0089	0179	0313	0500	0750	1071	1473	1964	2554	3250
16	0007	0029	0074	0147	0257	0412	0618	0882	1213	1618	2103	2676
17	0006	0025	0061	0123	0214	0343	0515	0735	1011	1348	1752	2230
18	0005	0021	0052	0103	0181	0289	0433	0619	0851	1135	1476	1878
19	0004	0018	0044	0088	0154	0246	0368	0526	0724	0965	1254	1596
20	0004	0015	0038	0075	0132	0211	0316	0451	0620	0827	1075	1368
21	0003	0013	0032	0065	0114	0182	0273	0390	0536	0714	0929	1182
22	0003	0011	0028	0056	0099	0158	0237	0339	0466	0621	0807	1028
23	0002	0010	0025	0049	0086	0138	0208	0296	0408	0543	0708	0899
24	0002	0009	0022	0043	0076	0122	0183	0261	0359	0478	0622	0791
25	0002	0008	0019	0038	0067	0108	0162	0231	0317	0423	0550	0700
26	0002	0007	0017	0034	0060	0096	0144	0205	0282	0376	0489	0622
27	0002	0006	0015	0031	0053	0085	0128	0183	0252	0336	0437	0556
28	0001	0005	0014	0027	0048	0077	0115	0164	0226	0301	0391	0498
29	0001	0005	0012	0025	0043	0069	0103	0148	0203	0271	0352	0448
30	0001	0004	0011	0022	0039	0062	0093	0133	0184	0245	0318	0405
35	0001	0003	0007	0014	0025	0039	0059	0084	0116	0154	0200	0255
40	0000	0002	0005	0009	0016	0026	0039	0056	0077	0103	0134	0171
45	0000	0001	0003	0007	0012	0018	0028	0040	0054	0072	0094	0120
50	0000	0001	0002	0004	0007	0011	0016	0023	0032	0043	0055	0070
60	0000	0001	0001	0003	0005	0008	0012	0017	0023	0031	0040	0051
70	0000	0000	0001	0002	0003	0005	0007	0010	0014	0019	0025	0032
80	0000	0000	0001	0001	0002	0003	0005	0007	0010	0013	0017	0021
90	0000	0000	0000	0001	0001	0002	0003	0005	0007	0009	0012	0015
100	0000	0000	0000	0000	0001	0002	0003	0004	0005	0007	0009	0011



As an example of the use of the table we will consider the following problem, [2, p. 56], with the ranks assigned as for the mid-rank method.

Subject	I	II
A	1	2.5
B	4	10
C	4	2.5
D	4	5
E	4	7
F	4	2.5
G	7	8
H	8	2.5
I	9.5	6
J	9.5	12
K	11	11
L	13	13
M	13	9
N	13	14

For the mid-rank method we have

$$\sum_{i=1}^{14} D_i^2 = 119.5, N = 14,$$

$$\rho_M = 1 - \frac{6(119.5)}{14(196 - 1)} = 0.7374.$$

Referring to the table we find that

$K_i$	$\Delta K_i$	$\delta_{NK_i}$
2	0.5	0.0011
3	2.0	0.0044
4	5.0	0.0110
5	10.0	0.0220
Total	17.5	0.0385

We know that  $\bar{\rho} = 1 - \frac{6(119.5 + 17.5)}{14(196 - 1)} = 0.6989$  and in terms of  $\delta_{NK_i}$

$$\bar{\rho} = 0.7374 - 0.0385 = 0.6989$$

The value given by DuBois for his method is 0.7511.

**Conclusion.** A method has been developed for the treatment of rank correlation where there are groups of equal variates. The method consists of applying a generally small correction to the value as ordinarily calculated by the mid-rank method in order to find the value which would be obtained by averaging the values of the rank correlation coefficient for all possible ways of arbitrarily assigning ranks to the equal variates. Thanks are due Professor P. S. Dwyer, without whose aid and encouragement this paper would not have been written.

#### REFERENCES

- [1] YULE AND KENDALL, *Introduction to the Theory of Statistics*, p. 248. London, 1937.
- [2] P. DuBois, "Formulas and Tables for Rank Correlation," *Psychol. Rec.*, Vol. 3(1939), pp. 45-56.

UNIVERSITY OF MICHIGAN,  
ANN ARBOR, MICHIGAN

# NOTE ON THEORETICAL AND OBSERVED DISTRIBUTIONS OF REPETITIVE OCCURRENCES

By P. S. OLMSTEAD

**1. A simple problem of repetitive occurrences.** Two questions which the engineer often desires to answer whenever he has a new type of apparatus or a new design of an old type of apparatus are: How many times will it perform its intended function without failure? and How many times will it fail to perform its intended function in a given length of time? To do this, he selects a number of what he believes to be identical units of the apparatus and gives each unit a performance test under a uniform test procedure. The number of satisfactory operations prior to the first observed failure to perform this operation is called a "run" and is a measure of the type desired for each unit.

If it is assumed that the probability of failure at any operation is a constant,  $q$ , and the probability of satisfactory operation is  $1 - q$  or  $p$ , then the mathematical probability of runs of 0, 1, 2, 3 . . . satisfactory operations for any unit are

$$(1) \quad q, pq, p^2q, p^3q, \dots$$

respectively.

Let  $x$  denote the number of satisfactory operations in any run. The mean value of  $x$ , say  $m_x$ , is given by

$$(2) \quad m_x = \frac{p}{q}.$$

The variance of  $x$  is

$$(3) \quad \sigma_x^2 = \frac{p}{q^2}.$$

The first step in practice is to determine whether there exists a constant probability,  $p$ , by means of the application of the operation of statistical control.<sup>1</sup> Expressions (1), (2), and (3) provide the necessary information for doing this. When a constant probability exists as evidenced by at least 25 consecutive samples of 4 units each the following practical procedure has been found to be satisfactory.

1. An estimate of  $p$  (or  $q$ ), the sole parameter of the distribution, can be obtained from the average length of run in the sample. If  $p$  is less than 0.6 and if the sample size is large, a reasonably good estimate of  $p$  can be obtained from the proportion of the sample having runs of zero length.

2. The probability of getting runs of length  $x$  or more is  $p^x$ . Thus, if a minimum (or maximum) value of the probability,  $p^x$ , is chosen, a maximum

<sup>1</sup> W. A. Shewhart, "Statistical Method from the Viewpoint of Quality Control," The Department of Agriculture Graduate School, Washington, 1939, Chapter I.

(or minimum) expected length of run can be computed for use as a criterion for looking for assignable causes of variation in the length of individual runs by using the estimated value of  $p$ .

3. The average and standard deviation to be used in calculating the limits to be applied to successive samples of rational sub-groups in accordance with the Shewhart<sup>2</sup> Criterion I are given by Equations (2) and (3) in which the estimates of  $p$  and  $q$  are substituted.

**2. Application to a signal transmission problem.** The theoretical solution given above is a direct answer to the first question at the head of this note.

TABLE I

*Observed distributions of runs of  $x$  occurrences of event  $E$  for various test periods of apparatus life*

No. of Occurrences per Period	Freq.	Test Period									
		1	2	3	4	5	6	7	8	11	15
$x$											
0	$n_0$	878	1519	961	723	541	407	343	266	160	77
1	$n_1$	77	226	207	206	171	148	129	97	70	35
2	$n_2$	2	31	44	55	68	46	52	39	37	27
3	$n_3$	1	3	8	18	15	19	13	22	19	10
4	$n_4$		2	1	2	—	6	5	5	7	3
5	$n_5$			—	1	1	3	1	1	5	2
6	$n_6$			1			1		—	1	2
7	$n_7$								1	—	—
8	$n_8$									2	1
Sample Size	$n$	958	1781	1222	1005	796	630	543	431	301	157

The second question is also of interest particularly when failure to perform an operation does not impair the apparatus unit for performance of additional operations. In cases of this type, the engineer often lets his test continue for test periods of particular lengths, measured in numbers of operations or sometimes in intervals of time (i.e., time intervals are often considered to be proportional to numbers of operations) and observes the number of failures during the test period for each unit. Thus, he may, after he has assured himself that control exists, arrange his data for each test period to show the frequency of occurrence of 0, 1, 2, 3, ... failures per unit.

Data of this type which are typical of those found in other studies made

<sup>2</sup> Loc. cit.

during the past two years are presented in Table I. These were obtained in a signal transmission study in which the data for successive periods were obtained

TABLE II  
*Comparison of observed and theoretical values of averages and variances for distributions of Table I*

Statistic or Parameter		Test Period									
		1	2	3	4	5	6	7	8	11	15
$\bar{q} = \frac{n_0}{n}$	observed	.916	.853	.786	.719	.679	.646	.632	.617	.532	.491
$\bar{x}$	observed	.098	.171	.269	.381	.448	.543	.537	.633	.917	1.026
$m_x = \frac{\bar{p}}{\bar{q}}$	theoretical*	.091	.172	.272	.390	.471	.548	.583	.620	.881	1.039
	observed	.091	.200	.343	.497	.556	.832	.760	1.075	1.783	1.921
$\sigma_x^2 = \frac{\bar{p}}{\bar{q}^2}$	theoretical*	.098	.202	.345	.542	.693	.848	.924	1.005	1.658	2.117

\* Based on assumption that  $\bar{q}$  is the true value of  $q$ .

TABLE III  
*Theoretical distributions corresponding to distributions of Table I calculated by using  $\bar{q} = \frac{n_0}{n}$  as the true value of  $q$*

No. of Occurrences per Period	Freq.	Test Period									
		1	2	3	4	5	6				15
$x$											
0	$n_0^*$	878.0	1519.0	961.0	723.0	541.0	407.0	343.0	266.0	160.0	77.0
1	$n_1$	73.3	233.5	205.3	202.8	173.3	144.1	126.4	101.9	74.9	39.2
2	$n_2$	6.1	32.9	43.8	56.9	55.5	51.0	46.6	39.0	35.1	20.0
3	$n_3$	.5	4.8	9.4	16.0	17.8	18.0	17.1	14.9	16.5	10.2
4	$n_4$	.1	.7	2.0	4.5	5.7	6.4	6.3	5.7	7.7	5.2
5	$n_5$		.1	.4	1.3	1.8	2.3	2.3	2.2	3.6	2.6
6	$n_6$			.1	.4	.6	.8	.9	.8	1.7	1.4
7	$n_7$				.1	.2	.3	.3	.3	.8	.7
8	$n_8$					.1	.1	.1	.1	.4	.3
9 or over	$n_{9-\infty}$								.1	.3	.4
Sample Size	$n^*$	958	1781	1222	1005	796	630	543	431	301	157

\* The observed values of  $n_0$  and  $n$  form the basis for the calculated distributions.

for separate units. Since each set of these data passed the scrutiny for control, there is justification for assuming that a statistical universe exists and that its functional form may be derived from the observed distribution. It was found

that these data were consistent with the assumption that, where the probability of non-occurrence of a failure on a unit in the test period was  $q$ , the probability of exactly  $x$  failures on a unit was  $p^x q$ . This set of mathematical probabilities is shown in (1) with  $q$  redefined to apply in this case to non-occurrence of a failure.

Observed and "Theoretical" values of the averages and variances for the observed distributions are shown in Table II. The basis for calculating the theoretical values was to take the ratio (designated  $\bar{q}$ ) of  $n_0$  to  $n$  for each distribution as the estimate of the true value,  $q$ . Distributions as shown in Table III

TABLE IV

*Test of fit of theoretical to observed distributions (Table III and Table I, respectively)*

	Test Period									
	1	2	3	4	5	6	7	8	11	15
	$\chi^2$ *	Degrees of Freedom	$P_{\chi^2}$							
	2.24	0.20	0.32	2.09	9.79	0.65	3.20	6.27	1.07	3.98
	1	2	2	3	3	3	3	3	4	4
	.13	.90	.87	.55	.02	.87	.36	.10	.90	.41

\* Minimum number in cell for theoretical distribution taken as 5.

were calculated from each  $\bar{q}$ . These distributions were tested against the observed distributions by means of the  $\chi^2$  test with the results shown in Table IV, which are all within reasonable limits of what might be expected when a constant probability exists.

**3. Conclusions.** When a constant probability applies to each operation in a repetitive process this note shows how to establish criteria for identifying significantly long or short lengths for individual runs and significantly high or low average lengths for groups of several runs. A problem taken from the field of signal transmission gives assurance of the existence of this type of distribution in practice.

BELL TELEPHONE LABORATORIES,  
NEW YORK, N. Y.

## THE DISTRIBUTION THEORY OF RUNS

By A. M. Mood

**1. Introduction.** In studying a particular sample, the order in which the elements of the sample were drawn is frequently available to the statistician. This important information is usually entirely neglected by him. Such disregard must be attributed, to a considerable extent, to the unsatisfactory state of mathematical devices for using the knowledge in question. One reasonable mathematical method for handling this information, the one to be used in this paper, is to make use of the distribution of runs. A run is defined as a succession of similar events preceded and succeeded by different events; the number of elements in a run will be referred to as its length.

The distribution theory of runs has had a stormy career. The theory seems to have been started toward the end of the nineteenth century rather than in the days of Laplace when there was so much interest in games of chance. In 1897 Karl Pearson [1], in a discussion of data taken from the roulette tables at Monte Carlo, wrote "... the theory of runs is a very simple one." In this book he developed no theory but it is evident from his computations that he regarded the distribution of runs as a special case of the multinomial distribution. The multinomial method, besides evading the issue somewhat and raising questions of random sampling, also gives incorrect results when one is interested in runs of more than one kind of element. In 1899 Karl Marbe [2] derived an expression for the mean of the number of iterations of a given length from a binomial population. This result was incorrect because he neglected dependence between overlapping iterations. An iteration is defined as a sequence of similar events; a run of length  $t$  is counted as  $t - s + 1$  iterations of length  $s$  for  $s \leq t$ . Marbe has assembled a great mass of data with the object of proving the popular hypothesis that a "head" becomes highly probable after a long succession of "tails" has appeared. Ordinary significance tests applied to his data do not support this contention, but Marbe continues to advocate it [3] and [5]. Of course, he has been severely criticised by many mathematical statisticians.

In 1904 Grünbaum [6] derived the mean of the number of runs of given length from a binomial population by the multinomial method. The first correct formulae were derived in 1906 by Bruns [7] who found the mean and variance of the number of iterations of given length in samples from a binomial population. In a book published in 1917 von Bortkiewicz correctly derived for the first time the mean and variance of runs from a binomial population using a method similar to that of Bruns. This book [8] contains a great many formulae for means and variances of runs and iterations under various special circumstances; a large portion of it is devoted to an exhaustive criticism of Marbe's work. In 1921 von

Mises [9] showed that the number of long runs of given length was approximately distributed according to the Poisson law for large samples.

It was not until 1925 (so far as the author has been able to ascertain) that an actual distribution function appeared when Ising [10] gave the number of ways of obtaining a given total number of runs (without regard to length) from arrangements of two kinds of elements. Stevens [12] in 1939 published the same distribution and described a  $\chi^2$  criterion for significance. Wald and Wolfowitz [13] in 1940 published the same distribution and showed that it was asymptotically normal. These papers are all concerned with random arrangements of a fixed number of elements of each of two kinds; the last mentioned paper describes a very interesting application of the distribution to the problem of testing the hypothesis that two samples have come from the same continuous distribution. Wishart and Hirshfeld [11] in 1936 derived the distribution of the total number of runs (again without regard to length) in samples from a binomial population and showed it was asymptotically normal.

In this paper we shall derive distributions of runs of given length both from random arrangements of fixed numbers of elements of two or more kinds, and from binomial and multinomial populations. Also we shall give the limiting form of these distributions as the sample size increases. These limiting distributions are all normal. The distribution problem is, of course, a combinatorial one, and the whole development depends on some identities in combinatory analysis,—some new and some well known to students of partition theory.

The paper will be divided into two parts. The first will deal with distributions obtained from random arrangements of a fixed number of each kind of element. The second will deal with distributions of elements from a binomial or multinomial population.

## PART I

**2. Distribution of runs of two kinds of elements.** Consider random arrangements of  $n$  elements of two kinds, for example  $n_1$   $a$ 's and  $n_2$   $b$ 's with  $n_1 + n_2 = n$ . Let  $r_{1i}$  denote the number of runs of  $a$ 's of length  $i$ , and let  $r_{2i}$  denote the number of runs of  $b$ 's of length  $i$ . For example the arrangement

$$a \ b \ b \ a \ b \ a \ a \ b \ b \ a \ a \ a$$

will be characterized by the numbers  $r_{11} = 2$ ,  $r_{13} = 2$ ,  $r_{21} = 1$ ,  $r_{22} = 2$ , and all other  $r_{ij} = 0$ . Also we let  $r_1 = \sum_i r_{1i}$  and  $r_2 = \sum_i r_{2i}$  denote the total number of runs of  $a$ 's and  $b$ 's respectively. Throughout the paper a binomial coefficient will be denoted by

$$(2.1) \qquad \binom{m}{k} = \frac{m!}{k!(m-k)!}$$

and this is defined to be zero when  $m < k$ . A multinomial coefficient will often be denoted by

$$(2.2) \quad \begin{bmatrix} m \\ m_i \end{bmatrix} = \frac{m!}{m_1! m_2! \dots m_s!}$$

$$(2.3) \quad \sum m_i = m, \quad m_i \geq 0$$

and when such a coefficient is to be summed over the indices  $m_i$  the two conditions (2.3) are always understood and will not be repeated; other conditions on the indices will be placed below the summation sign.

Given a set of numbers  $r_{ij}$  ( $i = 1, 2; j = 1, 2, \dots, n_i$ ) such that  $\sum_j j r_{ij} = n_i$ ,

there are  $\begin{bmatrix} r_1 \\ r_{1j} \end{bmatrix}$  and  $\begin{bmatrix} r_2 \\ r_{2j} \end{bmatrix}$  different arrangements of the runs of  $a$ 's and  $b$ 's respectively. Hence the total number of ways of obtaining the set  $r_{ij}$  is

$$(2.4) \quad N(r_{ij}) = \begin{bmatrix} r_1 \\ r_{1j} \end{bmatrix} \begin{bmatrix} r_2 \\ r_{2j} \end{bmatrix} F(r_1, r_2)$$

where  $F(r_1, r_2)$  is the number of ways of arranging  $r_1$  objects of one kind and  $r_2$  objects of another so that no two adjacent objects are of the same kind. Thus

$$(2.5) \quad \begin{aligned} F(r_1, r_2) &= 0 & \text{if } |r_1 - r_2| > 1, \\ &= 1 & \text{if } |r_1 - r_2| = 1, \\ &= 2 & \text{if } r_1 = r_2. \end{aligned}$$

Since there are  $\binom{n}{n_1}$  possible arrangements of the  $a$ 's and  $b$ 's, we have at once the distribution of the  $r_{ij}$

$$(2.6) \quad P(r_{ij}) = \frac{\begin{bmatrix} r_1 \\ r_{1j} \end{bmatrix} \begin{bmatrix} r_2 \\ r_{2j} \end{bmatrix} F(r_1, r_2)}{\binom{n}{n_1}}.$$

Certain marginal distributions will also be of interest. To obtain, for example, the distribution of the  $r_{1j}$ , it is first necessary to sum  $\begin{bmatrix} r_2 \\ r_{2j} \end{bmatrix}$  over all partitions of  $n_2$ . This is easily accomplished by finding the coefficient of  $x^{n_2}$  in

$$\begin{aligned} (x + x^2 + x^3 + \dots)^{r_2} &= x^{r_2} (1 + x + x^2 + \dots)^{r_2} = \frac{x^{r_2}}{(1-x)^{r_2}} \\ &= x^{r_2} \sum_{t=0}^{\infty} \binom{r_2-1+t}{r_2-1} x^t. \end{aligned}$$

The term corresponding to  $t = n_2 - r_2$  gives the desired result:

$$(2.7) \quad \sum_{\sum j r_{2j} = n_2} \begin{bmatrix} r_2 \\ r_{2j} \end{bmatrix} = \binom{n_2-1}{r_2-1}.$$



We have then

$$(2.8) \quad P(r_{1j}, r_2) = \frac{\begin{bmatrix} r_1 \\ r_{1j} \end{bmatrix} \binom{n_2 - 1}{r_2 - 1} F(r_1, r_2)}{\binom{n}{n_1}}$$

and summing this over  $r_2$ , a slight simplification gives

$$(2.9) \quad P(r_{1j}) = \frac{\begin{bmatrix} r_1 \\ r_{1j} \end{bmatrix} \binom{n_2 + 1}{r_1}}{\binom{n}{n_1}}.$$

The distribution (2.6) summed over  $r_{1j}$  and  $r_{2j}$  gives by means of (2.7)

$$(2.10) \quad P(r_1, r_2) = \frac{\binom{n_1 - 1}{r_1 - 1} \binom{n_2 - 1}{r_2 - 1} F(r_1, r_2)}{\binom{n}{n_1}}$$

which is essentially the distribution derived by Wald and Wolfowitz [13], and summing this over  $r_2$  we get the distribution discussed by Stevens [12]

$$(2.11) \quad P(r_1) = \frac{\binom{n_1 - 1}{r_1 - 1} \binom{n_2 + 1}{r_1}}{\binom{n}{n_1}}.$$

Another marginal distribution which will be useful is obtained by summing (2.9) over  $r_{1i}$  for  $i \geq k$ . If we let

$$\begin{aligned} s_{1j} &= r_{1j}, & j < k, \\ s_{1k} &= \sum_k^{n_1} r_{1j}, & A = \sum_1^{k-1} j r_{1j}, \end{aligned}$$

we must then sum the multinomial coefficient

$$\frac{s_{1k}!}{r_{1k}! \cdots r_{1n_1}!}$$

over all partitions of  $n_1 - A$  such that every part is greater than  $k - 1$ . This is given by the coefficient of  $x^{n_1 - A}$  in

$$(x^k + x^{k+1} + \cdots)^{s_{1k}} = x^{ks_{1k}} \sum_{t=0}^{\infty} \binom{s_{1k} - 1 + t}{s_{1k} - 1} x^t$$

thus we have

$$(2.12) \quad \sum_{(k)} \frac{s_{1k}!}{r_{1k}! \cdots r_{1n_1}!} = \binom{n_1 - A - (k - 1)s_{1k} - 1}{s_{1k} - 1}$$

where  $\sum_{(k)}$  denotes summation over all positive integers  $r_{1k}, r_{1k+1}, \dots, r_{1n_1}$  such that  $\sum_k jr_{1j} = n_1 - A$ . This identity with (2.9) gives

$$(2.13) \quad P(s_{1i}) = \frac{\begin{bmatrix} s_1 \\ s_{1i} \end{bmatrix} \binom{n_2 + 1}{s_1} \binom{n_1 - A - (k-1)s_{1k} - 1}{s_{1k} - 1}}{\binom{n}{n_1}}, \quad i = 1, 2, \dots, k.$$

Another useful distribution analogous to (2.13) is derived by considering runs of both kinds of elements. If we define  $s_{2j}$  ( $j = 1, 2, \dots, h$ ) and  $B$  in terms of  $r_{2j}$  just as  $s_{1i}$  and  $A$  were defined above, it follows at once from (2.6) and (2.12) that

$$(2.14) \quad P(s_{1i}, s_{2j}) = \frac{\begin{bmatrix} s_1 \\ s_{1i} \end{bmatrix} \begin{bmatrix} s_2 \\ s_{2j} \end{bmatrix} \binom{n_1 - A - (k-1)s_{1k} - 1}{s_{1k} - 1} \cdot \binom{n_2 - B - (h-1)s_{2h} - 1}{s_{2h} - 1} F(s_1, s_2)}{\binom{n}{n_1}} \\ i = 1, 2, \dots, k; j = 1, 2, \dots, h.$$

These last two distributions should be the most useful for applications. The long runs have been added together to form the new variables  $s_{1k}$  and  $s_{2h}$  thus decreasing materially the number of variables as compared with (2.6) and (2.9) while at the same time little information is lost. One is free to choose  $k$  and  $h$  so that the number of variables is appropriate for the data at hand. Moreover, it is shown in Section 5 that these variables are asymptotically normally distributed so that one may apply a simple  $\chi^2$  test of significance for "randomness of elements with respect to order" when dealing with large samples. We shall then be able to test whether a sample has been "randomly" drawn in a certain sense.

**3. Moments for runs of two kinds of elements.** Instead of dealing with the ordinary moments we shall obtain formulae for the factorial moments because the expressions are much more compact. As is customary, a factorial will be denoted by

$$(3.1) \quad x^{(a)} = x(x-1)(x-2) \dots (x-a+1),$$

and  $x^{(0)}$  is defined to be 1. Of course the ordinary moments are determined by the factorial moments by means of relations of the type

$$x^a = \sum_{i=0}^a C_i^a x^{(i)}.$$

A recent discussion of the coefficients  $C_i^a$  has been given by Joseph [14]. The mathematical expectation of a function  $f(r)$  will be denoted by

$$(3.2) \quad E(f(r)) = \sum_r f(r) P(r).$$

Of course  $E$  is a linear operator. We shall require the following identity

$$(3.3) \quad \sum_{(1)} \prod_i r_{1i}^{(a_i)} \begin{bmatrix} r_1 \\ r_{1i} \end{bmatrix} = r_1^{(\sum a_i)} \binom{n_1 - \sum ia_i - 1}{r_1 - \sum a_i - 1}$$

where  $\sum_{(1)}$  denotes summation over all positive integers  $r_{11}, r_{12}, \dots, r_{1n_1}$  such that  $\sum_1^{n_1} ir_{1i} = n_1$ . (3.3) may be verified by differentiating

$$\varphi(t_i) = (t_1x + t_2x^2 + \dots)^{n_1}$$

$a_i$  times with respect to  $t_i$  ( $i = 1, 2, \dots, n_1$ ), then finding the coefficient of  $x^{n_1}$  after putting  $t_i = 1$ . The identity (3.3) enables us to find the factorial moments of the variables in the distribution (2.9) for we have

$$\begin{aligned} E\left(\prod_i r_{1i}^{(a_i)}\right) &= \sum_{r_{1i}} \prod_i r_{1i}^{(a_i)} \begin{bmatrix} r_1 \\ r_{1i} \end{bmatrix} \binom{n_2 + 1}{r_1} / \binom{n}{n_1} \\ &= \sum r_1^{(\sum a_i)} \binom{n_1 - \sum ia_i - 1}{r_1 - \sum a_i - 1} \binom{n_2 + 1}{r_1} / \binom{n}{n_1} \\ (3.4) \quad &= \sum (n_2 + 1)^{(\sum a_i)} \binom{n_1 - \sum ia_i - 1}{r_1 - \sum a_i - 1} \binom{n_2 - \sum a_i + 1}{r_1 - \sum a_i} / \binom{n}{n_1} \\ &= (n_2 + 1)^{(\sum a_i)} \binom{n - \sum (i+1)a_i}{n_1 - \sum ia_i} / \binom{n}{n_1}. \end{aligned}$$

The sum on  $r_1$  involved in the last step is given by the identity

$$(3.5) \quad \sum_{i=0}^B \binom{A}{C+i} \binom{B}{i} = \binom{A+B}{C+B}$$

which is readily obtained by equating coefficients of  $x^C$  in

$$(1+x)^A \left(1 + \frac{1}{x}\right)^B = \frac{(1+x)^{A+B}}{x^B}.$$

We shall give here the means, variances and covariances obtained from (3.4)

$$(3.6) \quad E(r_{1i}) = (n_2 + 1)^{(2)} n_1^{(i)} / n^{(i+1)},$$

$$(3.7) \quad \sigma_{ii} = \frac{n_2^{(2)} (n_2 + 1)^{(2)} n_1^{(i+j)}}{n^{(i+j+2)}} - \frac{n_2^2 (n_2 + 1)^2 n_1^{(i)} n_1^{(j)}}{n^{(i+1)} n^{(j+1)}}$$

$$(3.8) \quad \sigma_{ii} = \frac{n_2^{(2)} (n_2 + 1)^{(2)} n_1^{(2i)}}{n^{(2i+2)}} + \frac{(n_2 + 1)^{(2)} n_1^{(i)}}{n^{(i+1)}} \left(1 - \frac{(n_2 + 1)^{(2)} n_1^{(i)}}{n^{(i+1)}}\right).$$

These will be needed in the section dealing with asymptotic distributions. The moments for the distribution (2.6) follow at once from (3.3) as

$$(3.9) \quad E \left( \prod_{i,j} r_{1i}^{(a_i)} r_{2j}^{(b_j)} \right) = \sum_{r_1, r_2} r_1^{(\sum a_i)} r_2^{(\sum b_j)} \cdot \binom{n_1 - \sum a_i - 1}{r_1 - \sum a_i - 1} \binom{n_2 - \sum b_j - 1}{r_2 - \sum b_j - 1} F(r_1, r_2) / \binom{n}{n_1}.$$

The summation on  $r_2$  is accomplished by putting  $r_2 = r_1 - 1$ ,  $r_1$ , and  $r_1 + 1$ , but after that has been done it is necessary to expand the product of the two factorial factors in factorial powers of the lower index of one of the binomial coefficients. This is easily done for the first few moments, but there appears to be no simple expression for the general case. The means, variances and covariances of  $r_1$ , are given by (3.6), (3.7) and (3.8) and those of  $r_2$ , are obtained from these equations by interchanging  $n_1$  and  $n_2$ . The other covariances are

$$(3.10) \quad \sigma_{r_1 r_2 j} = \frac{n_1^{(i+2)} n_2^{(j+2)}}{n^{(i+j+2)}} + 4 \frac{n_1^{(i+1)} n_2^{(j+1)}}{n^{(i+j+1)}} + 2 \frac{n_1^{(i)} n_2^{(j)}}{n^{(i+j)}} - \frac{(n_1 + 1)^{(2)} (n_2 + 1)^{(2)} n_1^{(i)} n_2^{(j)}}{n^{(i+1)} n^{(j+1)}}.$$

A slight variation of the method above will give the moments of the  $s_i$ ; in the distribution (2.13). An accent on a summation sign will indicate that the term corresponding to  $i = k$  is to be omitted. Differentiating

$$\varphi(t_i) = [t_1 x + t_2 x^2 + \dots + t_{k-1} x^{k-1} + t_k (x^k + x^{k+1} + \dots)]^{n_1}$$

$a_i$  times with respect to  $t_i$  and finding the coefficient of  $x^{n_1}$  after putting  $t_i = 1$ , we obtain

$$(3.11) \quad \sum_{i=1}^k \prod_{i=1}^k s_{1i}^{(a_i)} \left[ \frac{s_1}{s_{1k}} \right] \binom{n_1 - A - (k-1)s_{1k} - 1}{s_{1k} - 1} = s_1^{(\sum a_i)} \binom{n_1 - \sum a_i + a_k - 1}{s_1 - \sum' a_i - 1}.$$

This with (2.13) gives by the same steps as used in obtaining (3.4)

$$(3.12) \quad E \left( \prod_{i=1}^k s_{1i}^{(a_i)} \right) = (n_2 + 1)^{(\sum a_i)} \binom{n - \sum a_i - \sum' a_i}{n_1 - \sum a_i} / \binom{n}{n_1}.$$

The first two moments are

$$(3.13) \quad E(s_{1k}) = \frac{(n_2 + 1)n_1^{(k)}}{n^{(k)}},$$

$$(3.14) \quad \sigma_{ik} = \frac{n_2^2 (n_2 + 1) n_1^{(i+k)}}{n^{(i+k+1)}} - \frac{n_2 (n_2 + 1)^2 n_1^{(i)} n_1^{(k)}}{n^{(i+1)} n^{(k)}}.$$

$$(3.15) \quad \sigma_{kk} = \frac{(n_2 + 1)^{(2)} n_1^{(2k)}}{n^{(2k)}} + \frac{(n_2 + 1) n_1^{(k)}}{n^{(k)}} \left( 1 - \frac{(n_2 + 1) n_1^{(k)}}{n^{(k)}} \right).$$

The others are, of course, given by (3.6), (3.7) and (3.8).

The joint moments of the variables in (2.14) as obtained from (3.11) are

$$(3.16) \quad E \left( \prod_{ij} s_{1i}^{(a_i)} s_{2j}^{(b_j)} \right) = \sum_{s_1, s_2} s_1^{(\sum a_i)} s_2^{(\sum b_j)} \binom{n_1 - \sum i a_i + a_k - 1}{s_1 - \sum' a_i - 1} \cdot \binom{n_2 - \sum j b_j + b_k - 1}{s_2 - \sum' b_j - 1} F(s_1, s_2) / \binom{n}{n_1}.$$

In addition to the covariances (3.10) we shall need

$$(3.17) \quad \sigma_{s_1 k s_2 j} = \frac{n_1^{(k+2)} n_2^{(j+1)} + 2 n_1^{(k+1)} n_2^{(j+1)}}{n^{(k+j+1)}} + 2 \frac{n_1^{(k+1)} n_2^{(j)} + n_1^{(k)} n_2^{(j)}}{n^{(k+j)}} - \frac{(n_1 + 1)^{(2)} (n_2 + 1)^{(2)} n_1^{(k)} n_2^{(j)}}{n^{(k)} n^{(j+1)}},$$

$$(3.18) \quad \sigma_{s_1 k s_2 h} = \frac{n_1^{(k+1)} n_2^{(h+1)}}{n^{(k+h)}} + 2 \frac{n_1^{(k)} n_2^{(h)}}{n^{(k+h-1)}} - \frac{(n_1 + 1) (n_2 + 1) n_1^{(k)} n_2^{(h)}}{n^{(k)} n^{(h)}}.$$

The moments of  $r$  in the distribution (2.11) may be derived easily by means of (3.5) as

$$(3.19) \quad E(r_1^{(a)}) = (n_2 + 1)^{(a)} \binom{n - a}{n_1 - a} / \binom{n}{n_1}.$$

From which

$$(3.20) \quad E(r_1) = \frac{(n_2 + 1) n_1}{n},$$

$$(3.21) \quad \sigma_{r_1}^2 = \frac{(n_2 + 1)^{(2)} n_1^{(2)}}{n n^{(2)}}.$$

**4. Distribution and moments of runs of  $k$  kinds of elements.** This section is a generalization of the preceeding two sections to several kinds of elements. The case  $k = 2$  was treated separately because the special character of the function  $F(r_1, r_2)$  in this instance made the distribution comparatively simple. Now we shall be interested in  $k$  kinds of elements denoted by  $a_1, \dots, a_k$  and we shall suppose there are  $n_i$  elements of the  $i$ th kind. We let  $r_{ij}$  denote the number of runs of elements of the  $i$ th kind of length  $j$ , and put

$$n = \sum_1^k n_i, \quad r_i = \sum_{j=1}^{n_i} r_{ij}.$$

The same argument as was used in deriving (2.6) gives

$$(4.1) \quad P(r_{ij}) = \frac{\prod_{i=1}^k \left[ \begin{smallmatrix} r_i \\ r_{ij} \end{smallmatrix} \right] F(r_1, r_2, \dots, r_k)}{\left[ \begin{smallmatrix} n \\ n_i \end{smallmatrix} \right]}$$

where the function  $F(r_1, r_2, \dots, r_k)$ , which will be referred to hereafter simply as  $F(r_i)$ , represents the number of different arrangements of  $r_1$  objects of one kind,  $r_2$  objects of a second kind, and so forth, such that no two adjacent objects are of the same kind. We shall be able to give the explicit expression for  $F(r_i)$  after examining the marginal distribution  $P(r_i)$ . This is obtained by summing (4.1) over  $r_i$  with  $r_{ij}$  fixed by means of the identity (2.7) giving

$$(4.2) \quad P(r_i) = \frac{\prod_{i=1}^k \binom{n_i - 1}{r_i - 1} F(r_i)}{\left[ \begin{smallmatrix} n \\ n_i \end{smallmatrix} \right]}.$$

Despite our present meager knowledge of  $F(r_i)$  it is possible to find the moments of the  $r_i$  as distributed by (4.2). Since  $\sum_{r_i} P(r_i) = 1$ , we have the identity

$$(4.3) \quad \sum_{r_i} \Pi \binom{n_i - 1}{r_i - 1} F(r_i) = \left[ \begin{smallmatrix} n \\ n_i \end{smallmatrix} \right].$$

From this the moments are easily derived. If we put

$$(4.4) \quad u_i = n_i - r_i$$

we have

$$\begin{aligned} \sum_{r_i} \Pi u_i^{(a_i)} \Pi \binom{n_i - 1}{r_i - 1} F(r_i) &= \sum_{r_i} \Pi (n_i - r_i)^{(a_i)} \Pi \binom{n_i - 1}{r_i - 1} F(r_i) \\ &= \sum_{r_i} \Pi (n_i - 1)^{(a_i)} \Pi \binom{n_i - a_i - 1}{r_i - 1} F(r_i) \\ &= \Pi (n_i - 1)^{(a_i)} \sum_{r_i} \Pi \binom{n_i - a_i - 1}{r_i - 1} F(r_i) \\ &= \prod_{i=1}^k (n_i - 1)^{(a_i)} \left[ \begin{smallmatrix} n - \sum a_i \\ n_i - a_i \end{smallmatrix} \right]. \end{aligned}$$

The summation involved in the last step is given by (4.3). On dividing the last equation by  $\left[ \begin{smallmatrix} n \\ n_i \end{smallmatrix} \right]$  we get the factorial moments of the  $u_i$

$$(4.5) \quad E \left( \prod_1^k u_i^{(a_i)} \right) = \prod_1^k (n_i - 1)^{(a_i)} \left[ \begin{smallmatrix} n - \sum a_i \\ n_i - a_i \end{smallmatrix} \right] / \left[ \begin{smallmatrix} n \\ n_i \end{smallmatrix} \right].$$

From these equations the moments of the  $r_i$  may be found; the means, variances and covariances are

$$(4.6) \quad E(r_i) = \frac{n_i(n - n_i + 1)}{n}$$

$$(4.7) \quad \sigma_{ij} = \frac{n_i^{(2)} n_j^{(2)}}{n n^{(2)}},$$

$$(4.8) \quad \sigma_{ii} = \frac{n_i^{(2)}(n - n_i + 1)^{(2)}}{n n^{(2)}}.$$

It is clear that

$$(4.9) \quad \varphi(t_i) = \text{Coefficient of } \prod_1^k x_i^{n_i} \text{ in } (x_1 + \dots + x_k)^k \prod_1^k (x_1 + \dots + x_{i-1} + t_i x_i + x_{i+1} + \dots + x_k)^{n_i-1} / \left[ \begin{matrix} n \\ n_i \end{matrix} \right]$$

is a generating function for the moments of the variables  $u_i$ . This generating function will enable us to find the exact expression for  $F(r_i)$  for we have

$$\begin{aligned} P(u_i = n_{ii}) &= \text{Coefficient of } \prod_1^k t_i^{n_{ii}} \text{ in } \varphi(t_i) \\ &= \sum_{\substack{\alpha_i, n_{ij} \\ \sum_i \alpha_i = n_i - n_{ii}}} \left[ \begin{matrix} k \\ \alpha_i \end{matrix} \right] \prod_{i=1}^k \left[ \begin{matrix} n_i - 1 \\ n_{ij} \end{matrix} \right] / \left[ \begin{matrix} n \\ n_i \end{matrix} \right]. \end{aligned}$$

Also

$$P(u_i) = \prod_1^k \binom{n_i - 1}{r_i - 1} F(r_i) / \left[ \begin{matrix} n \\ n_i \end{matrix} \right]$$

and equating the expressions on the right of the last two equations we have

$$(4.10) \quad F(r_i) = \frac{\sum_{\alpha_i, n_{ij}} \left[ \begin{matrix} k \\ \alpha_i \end{matrix} \right] \prod_{i=1}^k \left[ \begin{matrix} n_i - 1 \\ n_{ij} \end{matrix} \right]}{\prod_1^k \binom{n_i - 1}{r_i - 1}}$$

$$(4.11) \quad = \sum_{\substack{\alpha_i, n'_{ij} \\ \sum_i \alpha_i = r_i - \alpha_j}} \left[ \begin{matrix} k \\ \alpha_i \end{matrix} \right] \prod_{i=1}^k \left[ \begin{matrix} r_i - 1 \\ n'_{ij} \end{matrix} \right]$$

in which the prime on the  $n'_{ij}$  indicates that the indices corresponding to  $j = i$  are to be omitted; hence  $i$  takes all the values  $1, 2, \dots, k$  and  $j$  takes all values  $1, 2, \dots, k$  except  $i$  because the index  $n_{ii}$  has been cancelled with  $n_i - r_i$  in the binomial coefficient in the denominator of (4.10). It is clear from (4.11) that  $F(r_i)$  may be expressed as follows

$$(4.12) \quad F(r_i) = CT \prod_1^k x_i^{-r_i} (x_1 + \dots + x_k)^k (x_2 + x_3 + \dots + x_k)^{r_1-1} (x_1 + x_3 + \dots + x_k)^{r_2-1} \dots (x_1 + \dots + x_{k-1})^{r_{k-1}-1}$$

in which "CT" is an abbreviation for "constant term of."

We are now in a position to obtain moments of the variables  $r_{ij}$  in the distribution (4.1) by means of identities similar to (4.3). As an illustration we compute

$$\begin{aligned} \sum_{r_i} \binom{n_1 - a - 1}{r_1 - a - 1} \prod_2^k \binom{n_i - 1}{r_i - 1} F(r_i) &= \sum_{r_i} \binom{n_1 - a - 1}{r_1 - a - 1} \prod_2^k \binom{n_i - 1}{r_i - 1} \\ &\quad \cdot CT \prod_1^k x_i^{-r_i} \prod_1^k (x_1 + \dots + t_i x_i + \dots + x_k)^{r_i - 1} \Big]_{t_i=0} \\ &= CT \prod_1^k x_i^{-n_i} (x_1 + \dots + x_k)^{n-a} (x_2 + \dots + x_k)^a \\ &= \left[ \begin{matrix} n \\ n_i \end{matrix} \right] \frac{(n - n_i)^{(a)}}{n^{(a)}} \end{aligned}$$

or

$$(4.13) \quad \sum_{r_i} \binom{n_1 - a - 1}{r_1 - a - 1} \prod_2^k \binom{n_i - 1}{r_i - 1} F(r_i) = \frac{(n - a)! (n - n_i)^{(a)}}{\prod_1^k n_i!}.$$

The moments of  $r_{ij}$  may be computed from identities of this type together with (3.3). The first two moments are

$$(4.14) \quad E(r_{ij}) = (n - n_i + 1)^{(2)} n_i^{(j)} / n^{(j+t)}$$

$$(4.15) \quad E(r_{ij}^{(2)}) = n_i^{(2j)} (n - n_i)^{(2)} (n - n_i + 1)^{(2)} / n^{(2j+2)}$$

$$(4.16) \quad E(r_{ij} r_{it}) = n_i^{(j+t)} (n - n_i)^{(2)} (n - n_i + 1)^{(2)} / n^{(j+t+2)} \quad j \neq t$$

$$\begin{aligned} E(r_{ij} r_{it}) &= (n_i - j - 1)(n_s - t - 1) \frac{n_i^{(j-1)} n_s^{(t-1)}}{n^{(j+t+2)}} \{ (n_i - j + 1)^{(2)} (n_s - t + 1)^{(2)} \\ &\quad + 2(n - n_i - n_s)(n_i - j + 1)(n_s - t + 1)(n_s - t + n_i - j) \\ &\quad + (n - n_i - n_s)^{(2)} [(n_s - t + 1)^{(2)} + 2(n_i - j + 1)(n_s - t + 1) \\ &\quad + (n_i - j + 1)^{(2)}] + 2(n - n_i - n_s)^{(3)} (n_i - j + n_s - t + 2) \\ &\quad + (n - n_i - n_s)^{(4)} \} + 2(n_i - j - 1) \frac{n_i^{(j-1)} n_s^{(t-1)}}{n^{(j+t+1)}} \{ (n_i - j + 1) \\ &\quad \cdot (n_s - t + 1)^{(2)} + (n - n_i - n_s)[2(n_i - j + 1)(n_s - t + 1) \\ &\quad + (n_s - t + 1)^{(2)}] + (n - n_i - n_s)^{(2)} [2(n_s - t + 1) + (n_i - j + 1)] \\ &\quad + (n - n_i - n_s)^{(3)} \} + 2(n_s - t - 1) \frac{n_i^{(j-1)} n_s^{(t-1)}}{n^{(j+t+1)}} \{ (n_s - t + 1) \\ &\quad \cdot (n_i - j + 1)^{(2)} + (n - n_i - n_s)[2(n_i - j + 1)(n_s - t + 1) \\ &\quad + (n_i - j + 1)^{(2)}] + (n - n_i - n_s)^{(2)} [2(n_i - j + 1) + (n_s - t + 1)] \\ &\quad + (n - n_i - n_s)^{(3)} \} + 2 \frac{n_i^{(j-1)} n_s^{(t-1)}}{n^{(j+t)}} \{ (n_i - j + 1)(n_s - t + 1) \\ &\quad + (n - n_i - n_s)(n_i - j + n_s - t + 2) + (n - n_i - n_s)^{(2)} \}. \end{aligned} \quad (4.17)$$



Such a lengthy expression as this last one can hardly be useful to the statistician, and for this reason we shall not define variables  $s_{ij}$  analogous to the  $s_{1i}$  and  $s_{2j}$  of Section 2 and take the time and space to find their moments.

**5. Asymptotic distributions.** We shall show that some of the distributions obtained previously are asymptotically normal when the  $n_i$  become large in such a way that the ratios  $n_i/n$  remain fixed. The description "asymptotically normal" means that the distribution approaches the normal distribution uniformly over any finite region as  $n_i \rightarrow \infty$ . The ratios  $n_i/n$  will be denoted by  $e_i$ , hence  $\sum e_i = 1$ . The symbol  $O(1/n^a)$  will represent any function such that

$$\lim_{n \rightarrow \infty} n^a O\left(\frac{1}{n^a}\right) = L < \infty.$$

We shall not, of course, be able to get any limit theorems for distributions like (2.6) or (2.9) because the number of independent variables increases with  $n$ . We shall consider first the distribution (2.13) whose asymptotic character is given in the following theorem.

**THEOREM 1.** *The variables*

$$(5.1) \quad \begin{aligned} x_i &= \frac{s_{1i} - ne_1^i e_2^2}{\sqrt{n}} & i < k \\ x_k &= \frac{s_{1k} - ne_1^k e_2}{\sqrt{n}} \end{aligned}$$

*are asymptotically normally distributed with zero means and variances and covariances*

$$(5.2) \quad \begin{aligned} \sigma_{ij} &= e_1^{i+j-1} e_2^3 [(i+1)(j+1)e_1 e_2 - i j e_2 - 2e_1], \quad i, j < k, \quad i \neq j \\ \sigma_{ii} &= e_1^{2i-1} e_2^3 [(i+1)^2 e_1 e_2 - i^2 e_2 - 2e_1] + e_1^i e_2^2, \quad i < k \\ \sigma_{ik} &= e_1^{i+k-1} e_2^2 [(i+1)k e_1 e_2 - i k e_2 - e_1], \quad i < k \\ \sigma_{kk} &= e_1^{2k-1} e_2^2 [k^2 (e_1 - 1) e_2 - e_1] + e_1^k e_2. \end{aligned}$$

The limiting means, variances and covariances are obtained from the relations (3.6), (3.7), (3.8), (3.13), (3.14) and (3.15).

To demonstrate this theorem we make the substitutions

$$(5.3) \quad \begin{aligned} n_i &= ne_i & i &= 1, 2 \\ s_{1i} &= ne_1^i e_2^2 + \sqrt{n} x_i & i &= 1, 2, \dots, k-1 \\ s_{1k} &= ne_1^k e_2 + \sqrt{n} x_k \\ s_1 &= ne_1 e_2 + \sqrt{n} \sum_1^k x_i \\ A &= n(e_1 - e_1^k - k e_1^k e_2) + \sqrt{n} \sum_1^{k-1} i x_i \end{aligned}$$

in (2.13), and estimate the factorials by means of Stirling's formula

$$(5.4) \quad m! = \sqrt{2\pi m}^{m+1/2} e^{-m} \left(1 + O\left(\frac{1}{m}\right)\right).$$

The result is an unwieldy expression which we shall not present at the moment. First we note that the exponential factors cancel out because the sum of the lower indices of a binomial or multinomial coefficient is equal to the upper index. Also we simplify the expression by considering in detail only terms which involve the  $x_i$ ; the normalizing constant can be determined from the final limit function. Any function of the parameters will be represented by the letter  $K$ . Thus in (5.4) we need consider only the factor  $m^{m+1/2}$ . All factorials will be of the form

$$(5.5) \quad m = na + \sqrt{n}L(x) + b$$

where  $L(x)$  is a linear function of the  $x_i$ , and  $a$  and  $b$  are independent of  $n$  and  $x_i$ . Now

$$\begin{aligned} m^{m+1/2} &= (na + \sqrt{n}L(x) + b)^{na + \sqrt{n}L(x) + b + 1/2} \\ &= (na)^{na + \sqrt{n}L(x) + b + 1/2} \left(1 + \frac{L(x)}{a\sqrt{n}} + \frac{b}{an}\right)^{na + \sqrt{n}L(x) + b + 1/2} \\ &= K(na)^{\sqrt{n}L(x)} \left(1 + \frac{L(x)}{a\sqrt{n}} + \frac{b}{an}\right)^{na + \sqrt{n}L(x) + b + 1/2} \end{aligned}$$

$$\begin{aligned} \text{and } \log m^{m+1/2} &= K + \sqrt{n}L(x) \log na + (na + \frac{5}{2}\sqrt{n}L(x) + b + \frac{1}{2}) \\ &\quad \cdot \log \left(1 + \frac{L(x)}{a\sqrt{n}} + \frac{b}{an}\right) \\ (5.6) \quad &= K + \sqrt{n}L(x) \log na + (na + \sqrt{n}L(x) + b + \frac{1}{2}) \\ &\quad \cdot \left(\frac{L(x)}{a\sqrt{n}} + \frac{b}{an} - \frac{L^2(x)}{a^2n} + O\left(\frac{1}{n^{3/2}}\right)\right) \\ &= K + \sqrt{n}L(x)(1 + \log na) + \frac{1}{2a}L^2(x) + O\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

so terms arising from  $b$  (and  $b + \frac{1}{2}$  in the exponent) will be neglected as they give rise only to terms independent of the  $x_i$  or of order  $1/n^{1/2}$ . Of course  $\log(1 + O(1/m)) = O(1/m)$ . Thus, keeping significant terms only, the result of the substitutions (5.3) and (5.4) in (2.13) after taking logarithms and using (5.6) is

$$\begin{aligned} -\log P(r_i) &= K + \sqrt{n} \sum_1^{k-1} x_i (\log ne_1^i e_2^2 + 1) + \sum_1^{k-1} \frac{x_i^2}{2e_1^i e_2^3} \\ &\quad - \sqrt{n} \left(\sum_1^k x_i\right) (\log ne_2^2 + 1) + \frac{1}{2e_2^3} \left(\sum_1^k x_i\right)^2 \end{aligned}$$

$$\begin{aligned}
 (5.7) \quad & + \sqrt{n} \left( \sum_1^{k-1} ix_i + (k-1)x_k \right) (\log ne_1^k + 1) - \frac{1}{2e_1^k} \left( \sum_1^k ix_i + (k-1)x_k \right)^2 \\
 & + 2\sqrt{n}x_k (\log ne_1^k e_2 + 1) + \frac{x_k^2}{e_1^k e_2} - \sqrt{n} \left( \sum_1^k ix_i \right) (\log ne_1^{k+1} + 1) \\
 & + \frac{1}{2e_1^{k+1}} \left( \sum_1^k ix_i \right)^2 + O\left(\frac{1}{\sqrt{n}}\right).
 \end{aligned}$$

The coefficients of  $x_i (i < k)$  and  $x_k$  are

$$\begin{aligned}
 \sqrt{n}(\log ne_1^k e_2 + 1 - \log ne_2^2 - 1 + i \log ne_1^k + i - i \log ne_1^{k+1} - i) &= 0, \\
 \sqrt{n}(-\log ne_2^2 - 1 + k \log ne_1^k + k + 2 \log ne_1^k e_2 + 2 - k \log ne_1^{k+1} - k) &= 0.
 \end{aligned}$$

Hence only the quadratic terms remain and we have

$$(5.8) \quad -\log P = K + \frac{1}{2} \sum_{i,j} \sigma^{ij} x_i x_j + O\left(\frac{1}{\sqrt{n}}\right)$$

where

$$\begin{aligned}
 \sigma^{ij} &= \frac{1}{e_2^2} + \frac{ij e_2}{e_1^{k+1}} & i, j < k, i \neq j, \\
 \sigma^{ii} &= \frac{1}{e_2^2} + \frac{1}{e_1^i e_2^2} + \frac{i^2 e_2}{e_1^{k+1}} & i < k, \\
 \sigma^{ik} &= \frac{1}{e_2^2} + \frac{i + i(k-1)e_2}{e_1^{k+1}} & i < k, \\
 \sigma^{kk} &= \frac{1}{e_2^2} + \frac{2}{e_1^k e_2} + \frac{k^2}{e_1^{k+1}} - \frac{(k-1)^2}{e_1^k}.
 \end{aligned}
 \tag{5.9}$$

It is merely a matter of straightforward multiplication of the two matrices to verify that  $\|\sigma^{ij}\|$  is the inverse of  $\|\sigma_{ij}\|$ , hence is a positive definite matrix. The details of the verification will be omitted. We have then

$$(5.10) \quad P = K e^{-\frac{1}{2} \sum \sigma^{ij} x_i x_j} \left( 1 + O\left(\frac{1}{\sqrt{n}}\right) \right).$$

In this equation  $K$  must necessarily contain the factor  $\left(\frac{1}{\sqrt{n}}\right)^k$  because there are  $k+5$  factorials in the denominator and 5 in the numerator of (2.13). Since  $\Delta r_i = 1$ , this factor, in view of (5.1), may be replaced by  $\Pi \Delta x_i$ , so

$$(5.11) \quad P = K e^{-\frac{1}{2} \sum \sigma^{ij} x_i x_j} \Pi \Delta x_i \left( 1 + O\left(\frac{1}{\sqrt{n}}\right) \right).$$

If we restrict the  $x_i$  to any finite region  $R$  in the  $x$ -space, the function  $O(1/\sqrt{n})$  approaches zero uniformly as  $n \rightarrow \infty$ . Thus, if  $A_i < B_i$  are any positive

numbers such that the corresponding values of  $x_i$ , say  $a_i$  and  $b_i$ , obtained by substituting  $A_i$  and  $B_i$  for  $r_i$  in (5.1), determine a rectangular region  $R'(a_i < x_i < b_i)$ , which lies in  $R$  we have

$$(5.12) \quad \sum_{r_i=A_i}^{B_i} P(r_i) = \sum_{x_i=a_i}^{b_i} K e^{-\sum \sigma^{ij} x_i x_j} \Pi \Delta x_i \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) \\ \int_{R'} K e^{-\sum \sigma^{ij} x_i x_j} \Pi dx_i$$

by the definition of a definite integral and Riemann's fundamental theorem.

We have given some details of this proof in order that it may serve as a model for other theorems of a similar nature which will appear later, and for which a complete proof will not be given. Two immediate consequences of Theorem 1 will now be stated as corollaries.

COROLLARY 1. *The variable*

$$x = \frac{r - ne_1 e_2}{\sqrt{ne_1 e_2}}$$

where  $r$  is the total number of runs of one kind of element, is asymptotically normally distributed with zero mean and unit variance. The limiting mean and variance were computed from (3.20) and (3.21).

COROLLARY 2. *The variable  $Q = \sum \sigma^{ij} x_i x_j$  is asymptotically distributed according to the  $\chi^2$ -law with  $k$  degrees of freedom.*

In exactly the same manner in which Theorem 1 was deduced from (2.13), we may prove the following theorem corresponding to the distribution (2.14).

THEOREM 2. *The variables*

$$(5.13) \quad x_i = \frac{s_{1i} - ne_1^i e_2^2}{\sqrt{n}} \quad i < k, \\ x_h = \frac{s_{1h} - ne_1^h e_2}{\sqrt{n}}, \\ y_i = \frac{s_{2i} - ne_1^2 e_2^i}{\sqrt{n}} \quad i < h,$$

are asymptotically normally distributed with] zero means and variances and covariances

$$\sigma_{x_i x_j} = e_1^{i+j-1} e_2^3 [(i+1)(j+1)e_1 e_2 - i j e_2 - 2e_1] \quad i, j < k, \\ \sigma_{x_i x_i} = e_1^{2i-1} e_2^3 [(i+1)^2 e_1 e_2 - i^2 e_2 - 2e_1] + e_1^i e_2^2 \quad i < k, \\ \sigma_{x_i x_h} = e_1^{i+h-1} e_2^2 [(i+1)k e_1 e_2 - i k e_2 - e_1] \quad i < k, \\ \sigma_{x_h x_h} = e_1^{2h-1} e_2^2 [-k^2 e_2^2 - e_1] + e_1^h e_2,$$

$$\begin{aligned}
 (5.14) \quad \sigma_{y_i y_j} &= e_2^{i+j-1} e_1^3 [(i+1)(j+1)e_1 e_2 - i j e_1 - 2e_2] & i, j < h, \\
 \sigma_{y_i y_i} &= e_2^{2i-1} e_1^3 [(i+1)^2 e_1 e_2 - i^2 e_1 - 2e_2] + e_2^i e_1^2 & i < h, \\
 \sigma_{x_i y_j} &= e_1^{i+1} e_2^{j+1} [(i+1)(j+1)e_1 e_2 - 2i e_2 - 2j e_1 + 4e_1 e_2 + 2] & i < k, j < h, \\
 \sigma_{x_h y_j} &= e_1^{k+1} e_2^j [k(j+1)e_1 e_2 - 2(k-1)e_2 - (j-1)e_1 + 2e_1 e_2] & j < h.
 \end{aligned}$$

These limiting variances were computed from the variances and covariances given in Section 3. We have chosen the variable  $s_{2h}$  of (2.14) as the dependent variable. The proof of this theorem is omitted. From it the following corollaries are deduced immediately.

**COROLLARY 3.** If  $u_i = x_i$  and  $u_{k+i} = y_i$  of (5.13) and  $\|\sigma^{ij}\|$  ( $i, j = 1, 2, \dots, k+h-1$ ) denotes the inverse of (5.14), then the variable  $Q = \Sigma \sigma^{ij} u_i u_j$  is asymptotically distributed according to the  $\chi^2$ -law with  $k+h-1$  degrees of freedom.

**COROLLARY 4.** If  $s_i = s_{1i} + s_{2i}$  denotes the total number of runs of both kinds of elements of length  $i$ , and  $s_k$  the total number of runs of length greater than  $k-1$ , then the variables

$$\begin{aligned}
 (5.15) \quad x_i &= \frac{s_i - n(e_1^i e_2^2 + e_2^i e_1^2)}{\sqrt{n}} & i < k \\
 x_h &= \frac{s_h - n(e_1^k e_2 + e_2^k e_1)}{\sqrt{n}}
 \end{aligned}$$

are asymptotically normally distributed with zero means and variances and covariances

$$(5.16) \quad \sigma_{ij} = \sigma_{x_i x_j} + \sigma_{x_i y_j} + \sigma_{x_j y_i} + \sigma_{y_i y_j}.$$

We have put  $h = k$  in Theorem 2 to obtain this result. The terms on the right of (5.16) are defined by (5.14); terms which do not appear there may be found by interchanging  $e_1$  and  $e_2$  in one of the relations. For example  $\sigma_{y_k y_k}$  is given by interchanging  $e_1$  and  $e_2$  in the fourth equation of the set (5.14).

**COROLLARY 5.** The variable  $Q = \Sigma \sigma^{ij} x_i x_j$ , where the  $x_i$  are defined by (5.15) and  $\|\sigma^{ij}\|$  is the inverse of (5.16), is asymptotically distributed according to the  $\chi^2$ -law with  $k$  degrees of freedom.

**COROLLARY 6.** If  $s$  denotes the total number of runs of both kinds of elements, then the variable

$$x = \frac{s - 2ne_1 e_2}{2\sqrt{ne_1 e_2}}$$

is asymptotically normally distributed with zero mean and unit variance. This is the result derived by Wald and Wolfowitz [13].

**6. Asymptotic distributions for  $k$  kinds of elements.** We now investigate the asymptotic character of the distribution (4.2)

$$(6.1) \quad P(r_i) = \frac{\prod_{i=1}^k \binom{n_i-1}{r_i-1} F(r_i)}{\left[ \begin{matrix} n \\ n_i \end{matrix} \right]}$$

where  $r_i$  is the total number of runs of the  $i$ th kind of element.

THEOREM 1. *If  $k > 2$ , the variables*

$$(6.2) \quad x_i = \frac{r_i - ne_i(1 - e_i)}{\sqrt{n}}$$

*are asymptotically normally distributed with zero means and variances and covariances*

$$(6.3) \quad \sigma_{ij} = e_i^2 e_j^2, \quad \sigma_{ii} = e_i^2(1 - e_i)^2.$$

The restriction  $k > 2$  is made because in the case  $k = 2$  the correlation between the two variables approaches one, and the numbers  $\sigma_{ij}$  are all equal. The result may be called a degenerate normal distribution and might be included in the theorem in this sense; we have chosen to omit it because this case is better taken care of by Corollary 1 of the previous section.

The proof of this theorem will be simplified if in the moments (4.5) we replace the numbers  $n_i - 1$  by  $n_i$ . This substitution will not, of course, affect the limiting moments. Hence we consider the variables  $v_i$  with moments given by

$$(6.4) \quad E\left(\prod_1^k v_i^{(a_i)}\right) = \frac{\prod_1^k n_i^{(a_i)} \left[ \begin{matrix} n - \sum a_i \\ n_i - a_i \end{matrix} \right]}{\left[ \begin{matrix} n \\ n_i \end{matrix} \right]}$$

and shall show that

$$(6.5) \quad y_i = \frac{v_i - ne_i^2}{\sqrt{n}}$$

are asymptotically normally distributed with zero means and variances and covariance (6.3). It is possible to prove this statement by showing that the characteristic function (Fourier transform) obtained by substituting  $i\theta_i$  for  $t_i$  in the moment generating function

$$(6.6) \quad \varphi_n(t_i) = \text{Coef. of } \prod_1^k x_i^{n_i} \text{ in } \prod_1^k (x_1 + \dots + x_{i-1} + t_i x_i + x_{i+1} + \dots + x_k)^{n_i} / \left[ \begin{matrix} n \\ n_i \end{matrix} \right]$$

approaches

$$\varphi(\theta_i) = e^{-\frac{1}{2} \sum \sigma_{ij} \theta_i \theta_j}$$

as  $n \rightarrow \infty$ . This method is not appropriate for proving a similar theorem which appears in Part II, and we prefer to give here a demonstration that will suffice for both theorems.

In order to prove our theorem we consider the general term in the coefficient of  $\Pi x_i^{n_i}$  in (6.6)

$$(6.7) \quad C(m_{ij}) = \prod_{i=1}^k \left[ \begin{matrix} n_i \\ m_{ij} \end{matrix} \right] \Pi t_i^{m_{ii}} / \left[ \begin{matrix} n \\ n_i \end{matrix} \right]$$

in which

$$(6.8) \quad \sum_{i=1}^k m_{ij} = n_j$$

must be required as well as the usual restriction on indices of a multinomial coefficient,  $\sum_{j=1}^k m_{ij} = n_i$ . Therefore only  $(k-1)^2$  of the indices are independent.

Clearly  $m_{ii} = v_i$ . Now without concerning ourselves about the statistical significance of the variables  $m_{ij}$ , let us consider their distribution

$$(6.9) \quad D(m_{ij}) = \prod_{i=1}^k \left[ \begin{matrix} n_i \\ m_{ij} \end{matrix} \right] / \left[ \begin{matrix} n \\ n_i \end{matrix} \right]$$

in which the variables corresponding to the values  $i, j = 1, 2, \dots, k-1$  will be chosen as the independent ones. We shall now prove a theorem from which Theorem 1 follows immediately.

**THEOREM 2.** *The variables*

$$(6.10) \quad x_{ij} = \frac{m_{ij} - ne_i e_j}{\sqrt{n}} \quad i, j = 1, 2, \dots, k-1$$

*are asymptotically normally distributed with zero means and variances and covariances given by*

$$(6.11) \quad \begin{aligned} \sigma_{ij,pq} &= e_i e_j e_p e_q, \\ \sigma_{ij,ip} &= -e_i(1-e_i)e_j e_p, \\ \sigma_{ij,ij} &= e_i e_j(1-e_i)(1-e_j). \end{aligned}$$

First it is to be noted that the moments of the  $m_{ij}$  are easily obtained from the identity

$$(6.12) \quad \sum_{\Sigma_i m_{ij} = n_i} \prod_{i=1}^k \left[ \begin{matrix} n_i \\ m_{ij} \end{matrix} \right] = \left[ \begin{matrix} n \\ n_i \end{matrix} \right]$$

as follows

$$\begin{aligned} \sum \prod_{ij} m_{ij}^{(a_{ij})} \prod_i \left[ \begin{matrix} n_i \\ m_{ij} \end{matrix} \right] &= \sum \prod_i n_i^{(\Sigma_j a_{ij})} \prod_i \left[ \begin{matrix} n_i - \sum_j a_{ij} \\ m_{ij} - a_{ij} \end{matrix} \right] \\ &\quad \prod_i n_i^{(\Sigma_j a_{ij})} \frac{n - \sum_{ij} a_{ij}}{n_j - \sum_{ij} a_{ij}} \end{aligned}$$

and on dividing this last relation by  $\begin{bmatrix} n \\ n_i \end{bmatrix}$  we obtain

$$(6.13) \quad E\left(\prod_{i,j} m_{ij}^{(a_{ij})}\right) = \prod_i n_i^{(\sum_j a_{ij})} \prod_j n_i^{(\sum_i a_{ij})} / n^{(\sum_{i,j} a_{ij})}$$

from which the moments (6.11) and the means in (6.10) were computed.

The proof of the theorem is similar to that of Theorem 1 in Section 5. We make the substitutions

$$\begin{aligned} n_i &= ne_i, & m_{ki} &= n_i - \sum_{j=1}^{k-1} m_{ij}, \\ m_{ik} &= n_i - \sum_{j=1}^{k-1} m_{ij}, & m_{kk} &= 2n_k + \sum_{i,j=1}^{k-1} m_{ij} - n, \\ m_{ij} &= ne_ie_j + \sqrt{nx_{ij}}, \end{aligned}$$

in (6.9) and employ Stirling's formula exactly as before. The details are too similar to warrant repetition. The final result is

$$(6.14) \quad D(m_{ij}) = Ke^{-i\sum \sigma^{ij,pq} x_{ij} x_{pq}} \Pi dx_{ij} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right).$$

Where  $\|\sigma^{ij,pq}\|$  is the inverse of (6.11) and is defined by

$$\begin{aligned} \sigma^{ij,pq} &= \frac{1}{e_k^2}, & \sigma^{ij,ij} &= \frac{1}{e_k^2} + \frac{1}{e_ie_k} + \frac{1}{e_je_k} + \frac{1}{e_ie_j}, \\ \sigma^{ij,ip} &= \frac{1}{e_1e_k} + \frac{1}{e_k^2}, & \sigma^{ij,pj} &= \frac{1}{e_1e_k} + \frac{1}{e_k^2}. \end{aligned}$$

Theorem 1 is a corollary of Theorem 2. Also we may state these additional results:

**COROLLARY 1.** *If  $k$  ( $\geq 3$ ) kinds of elements are arranged at random and  $r$  denotes the total number of runs of all kinds of elements, then the variable*

$$x = \frac{r - n(1 - \sum e_i^2)}{\sqrt{n}}$$

*is asymptotically normally distributed with zero mean and variance*

$$\sigma^2 = \sum e_i^2 - 2\sum e_i^3 + (\sum e_i^2)^2$$

*where  $e_i$  is the proportion of elements of the  $i$ -th kind.*

**COROLLARY 2.** *The variable  $Q = \sum \sigma^{ij} x_{ij}$ , where the  $x_{ij}$  are defined by (6.2) and  $\|\sigma^{ij}\|$  is the inverse of (6.3), is asymptotically distributed according to the  $\chi^2$ -law with  $k$  degrees of freedom.*

As was mentioned in Section 4, we could define variables  $s_{ij}$  ( $i = 1, 2, \dots, k$  and  $j = 1, 2, \dots, h_i$ , the  $h_i$  being a set of  $k$  arbitrary integers) with a distribution similar to (2.14). If one worked through the details he would find, no



doubt, that these variables are asymptotically normal. The matrix of variances and covariances is so complicated, however, that such a theorem would hardly be useful to the statistician, and the author does not feel that it would be worthwhile to go through the long and tedious details merely for the sake of completeness.

## PART II

Instead of having the number of elements of each kind fixed, we now suppose that they are randomly drawn from a binomial or multinomial population. The numbers  $n_i$  thus become random variables subject only to the restriction that  $\sum n_i = n$ , the sample number. The development will be entirely analogous to that of Part I, and the same notation will be used. The probability associated with the  $i$ th kind of element will be denoted by  $p_i$ .

**7. Distributions and moments.** The major part of the derivation of the various distribution functions has already been done in Sections 2 and 3. With the distributions of these sections we need only employ the fundamental relation

$$(7.1) \quad P(X, Y) = P_1(X | Y)P_2(Y)$$

in order to obtain the distributions required here.  $X$  will represent the set of variables  $r_{ij}$  or  $r_i$ , and  $Y$  the variables  $n_i$ . For the binomial population  $P_2(Y)$  will be

$$(7.2) \quad P(n_1, n_2) = \binom{n}{n_1} p_1^{n_1} p_2^{n_2}.$$

Therefore we may write down at once the distributions

$$(7.3) \quad P(r_{ij}, n_i) = \begin{bmatrix} r_1 \\ r_{1j} \end{bmatrix} \begin{bmatrix} r_2 \\ r_{2j} \end{bmatrix} F(r_1, r_2) p_1^{n_1} p_2^{n_2},$$

$$(7.4) \quad P(r_{1i}, n_i) = \begin{bmatrix} r_1 \\ r_{1i} \end{bmatrix} \binom{n_2 + 1}{r_1} p_1^{n_1} p_2^{n_2},$$

$$(7.5) \quad P(r_1, n_i) = \binom{n_1 - 1}{r_1 - 1} \binom{n_2 + 1}{r_1} p_1^{n_1} p_2^{n_2},$$

$$(7.6) \quad P(s_{1i}, n_i) = \begin{bmatrix} s_1 \\ s_{1i} \end{bmatrix} \binom{n_1 - A - (k-1)s_{1k} - 1}{s_{1k} - 1} \binom{n_2 + 1}{s_1} p_1^{n_1} p_2^{n_2},$$

$$(7.7) \quad P(s_{1i}, s_{2j}, n_i) = \begin{bmatrix} s_1 \\ s_{1i} \end{bmatrix} \begin{bmatrix} s_2 \\ s_{2j} \end{bmatrix} \binom{n_1 - A - (k-1)s_{1k} - 1}{s_{1k} - 1} \cdot \binom{n_2 - B - (h-1)s_{2h} - 1}{s_{2h} - 1} F(s_1, s_2) p_1^{n_1} p_2^{n_2},$$

$$i = 1, \dots, k, j = 1, \dots, h,$$

corresponding to the distributions (2.6), (2.9), (2.11), (2.13) and (2.14) respectively. Of course there is some dependence among the arguments. In (7.4), for example,  $n_1$  is determined by  $\sum i r_{1i} = n_1$ , and  $n_2$  by  $n - n_1 = n_2$ . In the last three distributions one of the  $n_i$  is independent and one may sum these with respect to  $n_1$  from zero to  $n$  and obtain the distributions of the  $r$ 's alone. The results of such summations are quite cumbersome and in some cases can only be indicated, so we shall retain the  $n_i$  as relevant variables. This remark applies also to the multinomial distribution.

We shall obtain expressions for the joint moments of the variables in these distributions. It is clear that the moments in Section 3 will be of considerable aid; for, using the notation of (7.1), we have

$$(7.8) \quad E(f(X)g(Y)) = \sum_{XY} f(X)g(Y)P(X, Y) = \sum_Y g(Y)P_2(Y) \left[ \sum_X f(X)P_1(X/Y) \right]$$

and the sum in the bracket on the right has been computed in Section 3. It remains only for us to multiply the previous moments by  $g(Y)P_2(Y)$  and sum on  $Y$ . Corresponding to (3.4), (3.12), (3.9) and (3.19) we have

$$(7.9) \quad E\left(n_1^{(a)} \prod_1^{n_1} r_{1i}^{(a_i)}\right) = \sum_{n_1=0}^n n_1^{(a)} (n_2 + 1)^{(\sum a_i)} \binom{n - \sum ia_i - \sum a_i}{n_1 - \sum ia_i} p_1^{n_1} p_2^{n_2},$$

$$(7.10) \quad E\left(n_1^{(a)} \prod_1^k s_{1i}^{(a_i)}\right) = \sum_{n_1=0}^n n_1^{(a)} (n_2 + 1)^{(\sum a_i)} \binom{n - \sum ia_i - \sum' a_i}{n_1 - \sum ia_i} p_1^{n_1} p_2^{n_2},$$

$$(7.11) \quad E(n_1^{(a)} r_1^{(b)}) = \sum_{n_1=0}^n n_1^{(a)} (n_2 + 1)^{(b)} \binom{n - b}{n_1 - b} p_1^{n_1} p_2^{n_2},$$

$$(7.12) \quad E\left(n_1^{(a)} \prod_1^k s_{1i}^{(a_i)} \prod_1^h s_{2j}^{(b_j)}\right) = \sum_{n_1, s_1, s_2} n_1^{(a)} s_1^{(\sum a_i)} s_2^{(\sum b_j)} \binom{n_1 - \sum ia_i + a_k - 1}{s_1 - \sum' a_i - 1} \cdot \binom{n_2 - \sum j b_j + b_h - 1}{s_2 - \sum' b_j - 1} F(s_1, s_2) p_1^{n_1} p_1^{n_2},$$

for moments from (7.4), (7.6), (7.5) and (7.7) respectively. In order to perform the summations indicated in these last relations it is necessary to expand the factors multiplying the binomial coefficient in factorial powers of its lower index. That is, we must write

$$(7.13) \quad n_1^{(a)} (n_2 + 1)^{(b)} = \sum_{i=0}^{a+b} C_i(n, a, b) (n_1 - b)^{(i)}.$$

Again it is not possible to give a simple expression for the coefficients  $C_i(n, a, b)$  in general, but for the first few moments they present no difficulty. For example from (7.9)

$$\begin{aligned}
 E(n_1 r_{1i}) &= \sum_{n_1=0}^n n_1(n - n_1 + 1) \binom{n-i-1}{n_1-i} p_1^{n_1} p_2^{n_1} \\
 &= \sum_{n_1} [i(n - i + 1) + (n - 2i)(n_1 - i) + (n_1 - i)^{(2)}] \\
 &\quad \cdot \binom{n-i-1}{n_1-i} p_1^{n_1} p_2^{n_1} \\
 (7.14) \quad &= \sum_{n_1} \left[ i(n - i + 1) \binom{n-i-1}{n_1-i} + (n - 2i)(n - i - 1) \right. \\
 &\quad \cdot \left. \binom{n-i-2}{n_1-i-1} - (n - i - 1)^{(2)} \binom{n-i-3}{n_1-i-2} \right] p_1^{n_1} p_2^{n_1} \\
 &= [i(n - i + 1) + (n - 2i)(n - i - 1)p_1 - (n - i - 1)^{(2)} p_1^2] p_1^i p_2.
 \end{aligned}$$

We give below some means, variances and covariances which will be required later.

$$\begin{aligned}
 E(r_{1i}) &= p_1^i p_2 [(n - i - 1)p_2 + 2], \\
 E(s_{1k}) &= p_1^k [(n - k)p_2 + 1], \\
 \sigma_{r_{1i}, r_{1j}} &= p_1^{i+j} p_2^2 \{ (n - i - j)^{(2)} p_2^2 + (n - i - j)p_2(1 + 5p_1) + 6p_1^2 \\
 &\quad - [(n - i - 1)p_2 + 2][(n - j - 1)p_2 + 2] \}, \\
 \sigma_{r_{1i}, r_{1i}} &= p_1^{2i} p_2^2 \{ (n - 2i)^{(2)} p_2^2 + (n - 2i)p_2(1 + 5p_1) + 6p_1^2 \\
 &\quad - [(n - i - 1)p_2 + 2]^2 \} + p_1^i p_2 [(n - i - 1)p_2 + 2], \\
 (7.15) \quad \sigma_{r_{1i}, r_{2j}} &= p_1^i p_2^j \{ (n - i - j - 2)^{(2)} p_1^2 p_2^2 + 4(n - i - j - 1)p_1 p_2 + 2 \\
 &\quad - [(n - i - 1)p_2 + 2][(n - j - 1)p_1 + 2] \}, \\
 \sigma_{s_{1i}, s_{1k}} &= p_1^{i+k} p_2 \{ (n - i - k + 1)^{(2)} - 2(n - i - k)^{(2)} p_1 \\
 &\quad + (n - i - k - 1)^{(2)} p_1^2 - [(n - i - 1)p_2 + 2][(n - k)p_2 + 1] \}, \\
 \sigma_{s_{1k}, s_{1k}} &= p_1^{2k} \{ (n - 2k + 1)^{(2)} - 2(n - 2k)^{(2)} p_1 + (n - 2k)^{(2)} p_1^2 \\
 &\quad - [(n - k)p_2 + 1]^2 \} + p_1^k [(n - k)p_2 + 1], \\
 \sigma_{s_{1k}, s_{2j}} &= p_1^k p_2^j \{ (n - k - j - 2)^{(2)} p_1^2 p_2^2 + 2(n - k - j - 1)p_1(1 + p_2) \\
 &\quad + 2(1 + p_1) - p_1[(n - k)p_2 + 1][(n - j - 1)p_1 + 2] \}.
 \end{aligned}$$

In order to obtain the distribution of runs in samples from a multinomial population, we multiply the distributions of Section 4 by

$$(7.16) \quad P(n_i) = \binom{n}{n_i} \prod_1^k p_i^{n_i}.$$

Corresponding to (4.1) and (4.2) then, we have

$$(7.17) \quad P(r_{ij}, n_i) = \prod_{i=1}^k \left[ \frac{r_i}{r_{ij}} \right] F(r_i) \prod_1^k p_i^{n_i}$$

$$(7.18) \quad P(r_i, n_i) = \prod_1^k \binom{n_i - 1}{r_i - 1} F(r_i) \prod_1^k p_i^{n_i}.$$

In (7.17)  $r_{ij}$  is the number of runs of length  $j$  of elements with probability  $p_i$ . In (7.18)  $r_i$  is the total number of runs of elements with probability  $p_i$ . As before, we shall investigate in detail only the distribution (7.18). The moments of  $n_i$  and  $r_i$  follow at once from (7.8) and (4.5)

$$(7.19) \quad E\left(\prod_1^k (n_i^{(a_i)} u_i^{(b_i)})\right) = \sum_{n_i} \prod_1^k (n_i^{(a_i)} (n_i - 1)^{(b_i)}) \left[ \frac{n - \sum b_i}{n_i - b_i} \right] \prod_1^k p_i^{n_i}$$

where  $u_i = n_i - r_i$ . The means, variances and covariances of the  $r_i$  are

$$E(r_i) = np_i(1 - p_i) + p_i^2,$$

$$(7.20) \quad \sigma_{r_i r_j} = -np_i p_j (1 - 2p_i - 2p_j + 3p_i p_j) - p_i p_j (2p_i + 2p_j - 5p_i p_j);$$

$$\sigma_{r_i r_i} = np_i(1 - 4p_i + 6p_i^2 - 3p_i^3) + p_i^2(3 - 8p_i + 5p_i^2).$$

**8. Asymptotic distributions from binomial population.** We turn our attention first to the distribution (7.7) and state a theorem analogous to Theorem 2 of Section 5.

**THEOREM 1.** *The variables*

$$(8.1) \quad \begin{aligned} u_i &= x_i = \frac{s_{1i} - np_1^i p_2^2}{\sqrt{n}}, & i &= 1, \dots, k-1, \\ u_k &= x_k = \frac{s_{1k} - np_1^k p_2}{\sqrt{n}}, \\ u_{k+i} &= y_i = \frac{s_{2i} - np_1^2 p_2^i}{\sqrt{n}}, & i &= 1, \dots, h-1, \\ u_{k+h} &= z = \frac{n_1 - np_1}{\sqrt{n}}, \end{aligned}$$

are asymptotically normally distributed with zero means and variances and covariances

$$(8.2) \quad \begin{aligned} \sigma_{x_i x_i} &= p_1^i p_2^2 - (2i+1)p_1^{2i} p_2^4 + 2p_1^{2i+1} p_2^3, \\ \sigma_{x_i x_j} &= -(i+j+1)p_1^{i+j} p_2^4 + 2p_1^{i+j+1} p_2^3, \\ \sigma_{x_i x_k} &= -(i+k+1)p_1^{i+k} p_2^3 + p_1^{i+k+1} p_2^2, \\ \sigma_{x_h x_h} &= p_1^k p_2 - (2k+1)p_1^{2k} p_2^3, \\ \sigma_{y_i y_j} &= -(i+j+1)p_1^4 p_2^{i+j} + 2p_1^3 p_2^{i+j+1}, \\ \sigma_{y_i y_i} &= p_1^3 p_2^i - (2i+1)p_1^4 p_2^{2i} + 2p_1^3 p_2^{2i+1}, \\ \sigma_{x_i y_j} &= -(i+j+3)p_1^{i+2} p_2^{j+2} + 2p_1^{i+1} p_2^{j+1}, \\ \sigma_{x_h y_j} &= -(k+j+2)p_1^{k+2} p_2^{j+1} + p_1^{k+1} p_2^j (1 + p_2), \\ \sigma_{x_i z} &= ip_1^i p_2^3 + p_1^{i+1} p_2 (1 - 4p_2), \\ \sigma_{x_h z} &= (k+1)p_1^k p_2^2 - p_1^k (1 + p_2), \\ \sigma_{y_i z} &= ip_1^3 p_2^i + p_1 p_2^{i+1} (1 - 4p_1), \\ \sigma_{zz} &= p_1 p_2. \end{aligned}$$

We have taken  $s_{2h}$  and  $n_2$  to be the dependent variables of (7.7). The method of proof of this theorem is the same as that of Theorem 1 in Section 5, and will be omitted. As consequences of the theorem we have

COROLLARY 1. *The variable*

$$Q = \sum_1^{k+h} \sigma^{ij} u_i u_j$$

*is asymptotically distributed according to the  $\chi^2$ -law with  $k + h$  degrees of freedom.*

COROLLARY 2. *Any subset  $u_{i_1}, u_{i_2}, \dots, u_{i_m}$  of the variables (8.1) is asymptotically normally distributed with zero means and variances and covariances  $\|\sigma_{i_j i_k}\|$ , and*

$$Q = \sum_{j,k=1}^m \sigma^{i_j i_k} u_{i_j} u_{i_k}$$

*is asymptotically distributed according to the  $\chi^2$ -law with  $m$  degrees of freedom.  $\|\sigma^{i_j i_k}\|$  is the inverse of  $\|\sigma_{i_j i_k}\|$ .*

COROLLARY 3. *If  $s_i = s_{1i} + s_{2i}$  represents the total number of runs of length  $i$  of both kinds of elements, and  $s_k$  the number of runs of length greater than  $k - 1$ , then the variables*

$$(8.3) \quad \begin{aligned} x_i &= \frac{s_i - n(p_1^i p_2^i + p_1^i p_2^i)}{\sqrt{n}}, & i = 1, \dots, k-1, \\ x_k &= \frac{s_k - n(p_1^k p_2 + p_1 p_2^k)}{\sqrt{n}}, \end{aligned}$$

*are asymptotically normally distributed with zero means and variances and covariances*

$$(8.4) \quad \sigma_{ij} = \sigma_{x_i x_j} + \sigma_{x_i y_j} + \sigma_{x_i y_j} + \sigma_{y_i y_j}$$

*where the terms on the right of (8.4) are defined by (8.2). We have put  $h = k$  in Theorem 1 to obtain this result.*

COROLLARY 4. *The variable*

$$(8.5) \quad Q = \sum_1^k \sigma^{ij} x_i x_j$$

*where the  $x_i$  are defined by (8.3) and  $\|\sigma^{ij}\|$  is the inverse of (8.4), is asymptotically distributed according to the  $\chi^2$ -law with  $k$  degrees of freedom.*

COROLLARY 5. *If  $r$  denotes the total number of runs of both kinds of elements, then*

$$(8.6) \quad x = \frac{r - 2np_1 p_2}{2\sqrt{np_1 p_2(1 - 3p_1 p_2)}}$$

*is asymptotically normally distributed with zero mean and unit variance. This is the result obtained by Wishart and Hirshfeld [11].*

**9. Asymptotic distributions from the multinomial population.** In this section we assume  $k > 2$  to avoid degenerate distributions. Because of the function  $F(r_i)$  in (7.18) we do not investigate this distribution directly, but derive a more general asymptotic distribution as was done in Section 6. We consider the distribution

$$(9.1) \quad D(m_{ij}, n_i) = \prod_{i=1}^k \left( \begin{bmatrix} n_i \\ m_{ij} \end{bmatrix} p_i^{m_{ij}} \right)$$

corresponding to (6.9). This is derived from (7.19) in the same manner as (6.9) was from (4.5). As before, we have replaced the numbers  $n_i - 1$  in (7.19) by  $n_i$ , an unessential change as far as the asymptotic theory is concerned. We recall that

$$(9.2) \quad r_i = n_i - m_{ii}$$

hence we need only show that the variables on the right are asymptotically normally distributed in order to have the same result for the  $r_i$ . Corresponding to Theorem 2 of Section 6, we state

**THEOREM 1.** *The variables*

$$(9.3) \quad \begin{aligned} x_{ij} &= \frac{m_{ij} - np_i p_j}{\sqrt{n}} & i, j = 1, \dots, k-1, \\ x_i &= \frac{n_i - np_i}{\sqrt{n}} & i = 1, \dots, k-1 \end{aligned}$$

*are asymptotically normally distributed with zero means and variances and covariances*

$$(9.4) \quad \begin{aligned} \sigma_{ij, st} &= -3p_i p_j p_s p_t, & \sigma_{ij, ii} &= -3p_i^2 p_j p_t, \\ \sigma_{ii, st} &= -3p_i^2 p_s p_t, & \sigma_{ii, ij} &= p_i^2 p_j (1 - 3p_i), \\ \sigma_{ij, ij} &= p_i p_j (1 - 3p_i p_j), & \sigma_{ii, ii} &= p_i^2 (1 + 2p_i - 3p_i^2), \\ \sigma_{ii, jj} &= -3p_i^2 p_j^2, & \sigma_{ij, s} &= -2p_i p_j p_s, \\ \sigma_{ii, s} &= -2p_i^2 p_s, & \sigma_{ij, i} &= p_i p_j (1 - 2p_i), \\ \sigma_{ii, i} &= 2p_i^2 (1 - p_i), & \sigma_{i, j} &= -p_i p_j, \\ & & \sigma_{i, i} &= p_i (1 - p_i). \end{aligned}$$

In these relations the symbols are defined by

$$\sigma_{ij, st} = \sigma_{x_{ij} x_{st}}, \quad \sigma_{ij, s} = \sigma_{x_{ij} x_s}, \quad \sigma_{i, j} = \sigma_{x_i x_j}$$

and different literal subscripts represent different numerical subscripts. These moments have been computed by means of the identity (6.12). The proof of the theorem is like that of Theorem 2 of Section 6 and will be omitted. We can now give the limiting form of the distribution of the  $r_i$  in (7.18) as

COROLLARY 1. *The variables*

$$(9.5) \quad x_i = \frac{r_i - np_i(1 - p_i)}{\sqrt{n}} \quad i = 1, 2, \dots, k$$

are asymptotically normally distributed with zero means and variances and covariances

$$(9.6) \quad \begin{aligned} \sigma_{ii} &= p_i(1 - p_i) - 3p_i^2(1 - p_i)^2, \\ \sigma_{ij} &= -p_i p_j(1 - 2p_i - 2p_j + 3p_i p_j). \end{aligned}$$

These limiting moments follow at once from equations (7.20).

COROLLARY 2. *The variable*

$$Q = \sum_1^k \sigma^{ij} x_i x_j$$

where the  $x_i$  are defined by (9.5) and  $\|\sigma^{ij}\|$  is the inverse of (9.6), is asymptotically distributed according to the  $\chi^2$ -law with  $k$  degrees of freedom.

COROLLARY 3. *If  $r = \Sigma r_i$  denotes the total number of runs, then*

$$x = \frac{r - n(1 - \Sigma p_i^2)}{\sqrt{n}}$$

is asymptotically normally distributed with zero mean and variance

$$\sigma^2 = \Sigma p_i^2 + 2\Sigma p_i^3 - 3(\Sigma p_i^2)^2.$$

The author would like to record here his gratitude to Professor S. S. Wilks who suggested the problem and under whose direction this paper was written.

#### REFERENCES

- [1] KARL PEARSON, *The Chances of Death and Other Studies in Evolution*, London, 1897, Vol. I, Chap. 2.
- [2] KARL MARBE, *Naturphilosophische Untersuchungen zur Wahrscheinlichkeitslehre*, Leipzig, 1899.
- [3] KARL MARBE, *Die Gleichförmigkeit in der Welt*, München, 1916.
- [4] KARL MARBE, *Mathematische Bemerkungen*, München, 1916.
- [5] KARL MARBE, *Grundfragen der angewandten Wahrscheinlichkeitsrechnung*, München, 1934.
- [6] H. GRÜNBAUM, *Isolierte und reine Gruppen und die Marbe'sche Zahl "p"*, Würzburg, 1904.
- [7] H. BRUNS, *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*, Leipzig, 1906, p. 216.
- [8] L. v. BORTKIEWICZ, *Die Iterationen*, Berlin, 1917.
- [9] R. v. MISES, *Zeit. f. angew. Math. u. Mech.*, Vol. 1 (1921), p. 298.
- [10] E. ISING, *Zeit. f. Physik.*, Vol. 31 (1925), p. 253.
- [11] J. WISHAART AND H. O. HIRSHFELD, *London Math. Soc. Jour.*, Vol. 11 (1936), p. 227.
- [12] W. L. STEVENS, *Annals of Eugenics*, Vol. 9 (1939), p. 10.
- [13] A. WALD AND J. WOLFOWITZ, *Annals of Math. Stat.*, Vol. 11 (1940).
- [14] J. A. JOSEPH, *Annals of Math. Stat.*, Vol. 10 (1939), p. 293.

PRINCETON UNIVERSITY,  
PRINCETON, N. J.

# A GENERALIZATION OF THE LAW OF LARGE NUMBERS

BY HILDA GEIRINGER

It is well known that the law of large numbers can be established for dependent as well as for independent chance variables by using Tchebycheff's inequality [1] and assuming that the variance of the sum of the variables tends towards infinity less rapidly than  $n^2$ .

In recent years v. Mises has introduced the notion of *statistical functions* [2] and has shown that, under certain assumptions the law of large numbers is still valid if, instead of the arithmetic mean of the  $n$  observations  $x_1, \dots, x_n$  a statistical function of these observations is considered. For example in the very special case, where the  $n$  collectives which have been observed are *identical*  $k$ -valued *arithmetic* distributions with probabilities  $p_1, \dots, p_k$  corresponding to the attributes  $c_1, \dots, c_k$  and with observed relative frequencies  $n_1/n, \dots, n_k/n$  one obtains the result: It is to be expected for every  $\epsilon > 0$  with a probability  $P_n$  converging towards one as  $n \rightarrow \infty$ , that  $|f(n_1/n, \dots, n_k/n) - f(p_1, \dots, p_k)| < \epsilon$  under very general conditions concerning the function  $f$ .

In the present paper we shall generalize these new results so that they will apply also to collectives which are not independent.

**1. Lemma concerning alternatives.** Let us consider the  $n$ -dimensional *collective* consisting of a *sequence of  $n$  trials* and let us assume that the  $n$  trials are alternatives, i.e. for each trial there are only two possible results which we denote by "success," "failure," by "occurrence," "non-occurrence" or by "1," "0." The total result of the  $n$  trials is expressed by  $n$  numbers each equal to 0 or 1. Let  $v(x_1, x_2, \dots, x_n)$  be the probability of obtaining the result  $x_1$  at the first trial,  $x_2$  at the second one,  $\dots$ ,  $x_n$  at the last one ( $x_\nu = 0, 1$ ;  $\nu = 1, \dots, n$ ). In the same way we introduce  $v_{12}(x, y) = \sum_{x_3, \dots, x_n} v(x, y, x_3, \dots, x_n)$  and generally  $v_{\mu\nu}(x, y)$  as the probability that the  $\mu$ th result equals  $x$ , the  $\nu$ th equals  $y$ , ( $\mu \neq \nu$ ), and finally let  $v_\mu(x) = \sum_\nu v_{\mu\nu}(x, y)$  be the probability that the  $\mu$ th result equals  $x$ . In particular let us write

$$v_\mu(1) = p_\mu, \quad v_{\mu\nu}(1, 1) = p_{\mu\nu}, \quad (\mu, \nu = 1, \dots, n; \mu \neq \nu)$$

$p_\mu$  being the probability of success in the  $\mu$ th trial and  $p_{\mu\nu}$  the probability of simultaneous success both in the  $\mu$ th and  $\nu$ th trials.

The *variance*  $s_n^2$  of the sum  $(x_1 + \dots + x_n)$  is easily found:



$$\begin{aligned}
s_n^2 &= \text{Var}(x_1 + \dots + x_n) = \sum_{x_1, \dots, x_n} (x_1 + \dots + x_n - p_1 - \dots - p_n)^2 v(x_1, \dots, x_n) \\
&= \sum_{x_1, \dots, x_n} (x_1 - p_1)^2 v(x_1, \dots, x_n) + \dots \\
&\quad + 2 \sum_{x_1, \dots, x_n} (x_1 - p_1)(x_2 - p_2) v(x_1, \dots, x_n) + \dots \\
&= \sum_{x_1} (x_1 - p_1)^2 v_1(x_1) + \dots + 2 \sum_{x_1, x_2} (x_1 - p_1)(x_2 - p_2) v_{12}(x_1, x_2) + \dots \\
&= p_1(1 - p_1) + \dots + p_n(1 - p_n) + 2(p_{12} - p_1 p_2) + \dots + 2(p_{n-1, n} - p_{n-1} p_n).
\end{aligned}$$

Thus:

$$(1) \quad s_n^2 = \text{Var}(x_1 + \dots + x_n) = \sum_{\nu=1}^n p_\nu(1 - p_\nu) + 2 \sum_{\mu, \nu=1}^n (p_{\mu\nu} - p_\mu p_\nu).$$

The first sum on the right is  $\leq n/4$ ; the second one consists of  $N = \frac{1}{2}n(n-1)$  terms, therefore we cannot be sure that it tends toward zero after division by  $n^2$ .

Putting  $p_{\mu\nu} - p_\mu p_\nu = \alpha_{\mu\nu}^{(n)}$  we see immediately:

(a) A necessary and sufficient condition for  $\lim_{n \rightarrow \infty} s_n/n = 0$  is

$$(2) \quad \lim_{n \rightarrow \infty} 1/n^2 \sum_{\mu, \nu=1}^n \alpha_{\mu\nu}^{(n)} = 0.$$

Denoting by  $\sigma_\mu^2$  the variance of  $v_\mu(x)$  and by  $r_{\mu\nu}$  the correlation coefficient of  $v_{\mu\nu}(x, y)$  we have

$$\alpha_{\mu\nu}^{(n)} = p_{\mu\nu} - p_\mu p_\nu = r_{\mu\nu} \sigma_\mu \sigma_\nu.$$

We see that  $\alpha_{\mu\nu}^{(n)}$  takes values between  $-1/4$  and  $+1/4$  and our conditions (2) postulates that the sum of these positive and negative terms tends towards infinity less rapidly than  $n^2$ . As to the meaning of the signs of these terms we

see that a term  $\alpha_{\mu\nu}^{(n)}$  will be  $\geq 0$ , according as  $p_{\mu\nu}/p_\nu \geq p_\mu$ . This means: the fact that the  $\nu$ th event has presented itself makes the occurrence of the  $\mu$ th event either more probable; or it is without influence on it; or it makes it less probable. And we see that  $s_n/n$  tends toward zero, only if there is a certain "equalization" or "stabilization" of positive and negative mutual influence. If in particular for a pair of values  $\mu, \nu$ ,  $r_{\mu\nu} = +1$ , that is  $v_{\mu\nu}(0, 1) = v_{\mu\nu}(1, 0) = 0$ , the events must either both occur or both fail and  $p_\mu = p_\nu$ . If  $r_{\mu\nu} = -1$  we have  $v_{\mu\nu}(0, 0) = v_{\mu\nu}(1, 1) = 0$  the simultaneous occurrence is impossible and likewise the simultaneous failure, and  $p_\mu + p_\nu = 1$ . If we have  $p_{\mu\nu} = 0$  (case of mutually exclusive events) then  $p_\mu + p_\nu \leq 1$ .

Since  $s_n^2 \geq 0$  and  $\sum_{\nu=1}^n p_\nu(1 - p_\nu) = \sum_{\nu=1}^n \sigma_\nu^2 \leq n/4$  we conclude from (1) that

$\sum_{\mu, \nu=1}^n \alpha_{\mu\nu}^{(n)} \geq -n/8$  and we obtain the following simple *sufficient* condition for the validity of (2):

(b) Let us denote by  $m_n$  the number of all combinations  $\mu, \nu$  ( $\mu \leq n; \nu \leq n; \mu \neq \nu$ ), such that, however large  $n$  may be,  $\alpha_{\mu\nu}^{(n)} > \epsilon$ , where  $\epsilon$  is a given positive number; then  $\frac{1}{n^2} \sum_{\mu, \nu=1}^n \alpha_{\mu\nu}^{(n)}$  converges toward zero if  $\lim_{n \rightarrow \infty} m_n/n^2 = 0$ .

We have in fact

$$-\frac{n}{8} \leq \sum_{\mu, \nu=1}^n \alpha_{\mu\nu}^{(n)} \leq m_n + (N - m_n)\epsilon$$

and dividing by  $n^2$  we find that  $\frac{1}{n^2} \sum_{\mu, \nu=1}^n \alpha_{\mu\nu}^{(n)}$  is enclosed between  $-\frac{1}{8n}$  and  $m_n/n^2 + \frac{N - m_n}{n^2}$  which both tend toward zero. Roughly speaking this condition implies that for "almost all" combinations of indices  $\mu, \nu$ , the  $\alpha_{\mu\nu}^{(n)}$  converge toward "negative or vanishing correlation."

On the other hand the sum of all positive and negative terms in  $\sum_{\mu, \nu=1}^n \alpha_{\mu\nu}^{(n)}$  cannot become less than  $-n/8$ . Therefore, if "almost all" positive terms are supposed to tend towards zero it follows that also almost all negative terms tend toward zero. Thus we obtain the *sufficient* condition (c) which is neither more nor less general than (b):

(c) The sum  $\frac{1}{n^2} \sum_{\mu, \nu=1}^n \alpha_{\mu\nu}^{(n)}$  tends towards zero as  $n \rightarrow \infty$ , if "almost all" the individual terms  $\alpha_{\mu\nu}^{(n)} = p_{\mu\nu} - p_\mu p_\nu$  tend toward zero. Or more exactly, the sum in question tends toward zero if  $|\alpha_{\mu\nu}^{(n)}| \leq \epsilon$  for every  $\epsilon$  and sufficiently large  $n$  with the exception of  $\mu_n$  terms where  $\lim_{n \rightarrow \infty} \mu_n/n^2 = 0$ . That is "convergence towards independence" for almost all combinations  $\mu, \nu$  of indices. Let us, for example, assume that all the  $p_\nu$  are  $\neq 0$  and all the  $p_{\mu\nu} = 0$ , then all the  $\alpha_{\mu\nu}^{(n)}$  are certainly  $< 0$  and (b) is fulfilled; but it is easily seen (3) that in this case  $p_1 + p_2 + \dots + p_n \leq 1$ . Therefore all the products  $p_\mu p_\nu$  (with the possible exception of a finite number) tend toward zero, and (c) holds as well.

**2. Statistical functions.** Suppose  $n$  observations have given the results  $x_1, x_2, \dots, x_n$ . Let us assume for the sake of simplicity that they are all bounded between two real numbers  $A$  and  $B$ . To each real  $x$  corresponds the number  $n S_n(x)$  of observations with a result  $\leq x$ .  $S_n(x)$  is a monotone non-decreasing step function with  $n$  steps, each of height  $1/n$ ; however several steps may coincide at the same point. We have

$$(1) \quad S_n(x) = 0 \quad \text{if } x < A \quad \text{and} \quad S_n(x) = 1 \quad \text{if } x \geq B.$$

$S_n(x)$  is called by v. Mises the *partition* (Aufteilung) of the  $n$  observations.  $S_n(x)$  coincides with the well known cumulative frequency distribution if the attributes  $c_k$  ( $k = 1, \dots, k$ ) and the corresponding relative frequencies  $n_1/n, \dots, n_k/n$  are given.

A *statistical function* is a function of the  $x_1, x_2, \dots, x_n$  which depends only on  $S_n(x)$ , the partition of the  $n$  results. It will be denoted by  $f\{S_n(x)\}$ . If the  $c_x$  and the  $n_x/n$  are given then statistical function means simply "function of the relative frequencies" and it becomes a function of  $k$  variables. In  $f\{S_n(x)\}$  the partition  $S_n(x)$  takes the place of the independent variable. Such a statistical function has the following properties: (a) It is a symmetric function of the  $x_1, x_2, \dots, x_n$ . That is, it is independent of the succession of the  $n$  results. (b) It is "homogeneous" in the following sense: If instead of  $n$  observations we have  $nl$  observations and if at the same time each  $x_i$  is replaced by  $lx_i$ , then the statistical function is not changed.<sup>1</sup> Examples of statistical functions are the *moments*

$$\frac{1}{n} \sum_{r=1}^n x_r^r = \int x^r dS_n(x) = M_r^0$$

or, if  $M_1^0 = \alpha$ , the moments about the mean  $\alpha$ :

$$\frac{1}{n} \sum_{r=1}^n (x_r - \alpha)^r = \int (x - \alpha)^r dS_n(x) = M_r, \text{ etc.}$$

The independent variable in  $f\{S_n(x)\}$  is a partition; but in addition we shall define  $f\{P(x)\}$  where  $P(x)$  is a certain bounded distribution which is not necessarily a partition. A distribution  $P(x)$  is called bounded if

$$(1') \quad P(x) = 0 \quad \text{if } x < A \quad \text{and} \quad P(x) = 1 \quad \text{if } x \geq B.$$

If this is true for a sequence  $P_1(x), P_2(x), \dots$  with the same  $A$  and  $B$  then the sequence is called *uniformly bounded*. Let us now consider a bounded partition  $P(x)$  which in every point of continuity of  $P(x)$  is the limit as  $n \rightarrow \infty$  of a sequence of bounded partitions  $S_n(x)$ . As  $S_n(x)$  converges toward  $P(x)$ , if  $f\{S_n(x)\}$  converges towards a limit  $L$  which does not depend on the limiting process  $S_n(x) \rightarrow P(x)$  then that limit shall be denoted by  $f\{P(x)\}$ ; it will be called the *value of the statistical function at the "point"  $P(x)$*  and  $f\{S_n(x)\}$  will be called *continuous* at  $P(x)$ . The definition of continuity can be given also in the following way: Corresponding to every  $\epsilon > 0$  exists an  $\eta > 0$  such that

$$(2) \quad |f\{S_n(x)\} - f\{P(x)\}| < \epsilon$$

for all values of  $n$  and for every bounded  $S_n(x)$  such that at every point of continuity of  $P(x)$

$$(3) \quad |S_n(x) - P(x)| \leq \eta.$$

In this case  $f\{S_n(x)\}$  is called continuous at the point  $P(x)$ . Thus a statistical function is defined for bounded partitions and for certain bounded distributions which are not themselves partitions. If the continuity defined by (2) and (3) exists for a sequence  $P_1(x), P_2(x), \dots$  of bounded distributions with the same  $\eta$

<sup>1</sup> This condition of homogeneity is fulfilled e.g. for  $\sqrt{x_1 x_2 \dots x_n}$  but not for  $x_1 x_2 \dots x_n$ .

corresponding to a given  $\epsilon$ , we call the statistical function *uniformly* continuous at the points  $P_1(x), P_2(x), \dots$ .

**3. The general law of large numbers.** The generalization of the law of large numbers which we have in mind can be demonstrated in a way analogous to the demonstration given by v. Mises in the case of independent collectives if we introduce the results of paragraph 1 in order to estimate the variance. We shall consider here only one dimensional, bounded collectives in order to make clearer what is the essential of the generalization.

A sequence of dependent collectives  $P_1(x), P_2(x), \dots, P_n(x)$  can be given in the following manner. Let  $P(x_1, x_2, \dots, x_n)$  be the probability that the result of the first observation is  $\leq x_1$ , of the second  $\leq x_2, \dots$ , of the  $n$ th  $\leq x_n$ . This distribution will be said to be *bounded* in  $(A, B)$  if  $P = 1$  when all the  $x_i$  are  $\geq B$  and  $P = 0$  if at least one of these arguments is less than  $A$ . From this  $n$ -dimensional distribution we deduce  $n$  one dimensional distributions

$$(1) \quad \begin{aligned} P_1(x) &= P(x, B, \dots, B), \\ P_2(x) &= P(B, x, B, \dots, B), \dots, P_n(x) = P(B, \dots, B, x) \end{aligned}$$

where  $P_\nu(x)$  is the probability that the  $\nu$ th observation be  $\leq x$ . The  $P_\nu(x)$  are uniformly bounded in  $(A, B)$  which is a consequence of  $P(x_1, x_2, \dots, x_n)$  having been assumed to be bounded in this interval. In an analogous way we deduce from  $P(x_1, x_2, \dots, x_n)$  the  $\frac{1}{2}n(n-1)$  uniformly bounded two dimensional distributions

$$(2) \quad P_{12}(x, y) = P(x, y, B, \dots B), \quad P_{13}(x, y) = P(x, B, y, B, \dots B), \dots$$

Here  $P_{\mu\nu}(x, y)$  is the probability that the  $\mu$ th result is  $\leq x$ , the  $\nu$ th result  $\leq y$ , and we have  $P_{\mu\nu}(x, y) = P_{\nu\mu}(y, x)$ . Of course we have also

$$(1') \quad \begin{aligned} P_1(x) &= P_{12}(x, B) = P_{13}(x, B) = \dots = P_{1n}(x, B) \\ P_2(x) &= P_{12}(B, x) = P_{23}(x, B) = \dots = P_{2n}(x, B) \text{ etc.} \end{aligned}$$

If we put in (2)  $x = y$  we obtain  $P_{\mu\nu}(x, x) = P_{\nu\mu}(x, x)$  and we introduce

$$(3) \quad P_{\mu\nu}(x, x) = P_{\nu\mu}(x) = P_{\mu\nu}(x)$$

the probability that both the  $\mu$ th and the  $\nu$ th observation is  $\leq x$ . Then  $P_{\mu\nu}(x)$  equals zero if  $x < A$  and equals one if  $x \geq B$ , and this is valid with the same  $A$  and  $B$  for all the distributions  $P_{\mu\nu}(x)$ .

Now if  $p_1, p_2, \dots, p_n$  are the probabilities of success for  $n$  general alternatives Tchebycheff's Lemma asserts that the probability  $W$  that the average  $(x_1 + x_2 + \dots + x_n)/n$  of  $n$  observations differs by more than  $\eta$  from its expectation  $(p_1 + p_2 + \dots + p_n)/n$  is subject to the following inequality

$$(4) \quad W \leq \frac{1}{\eta^2} \text{Var} \left( \frac{x_1 + x_2 + \dots + x_n}{n} \right) = \frac{s_n^2}{\eta^2 n^2}.$$

Here  $s_n^2$  is given by (1) of paragraph 1.

Let us introduce the average  $\bar{P}_n(x)$  of the  $P_r(x)$ :

$$(5) \quad \bar{P}_n(x) = [P_1(x) + P_2(x) + \dots + P_n(x)]/n$$

and let  $Q_n$  be the probability that at any point of continuity of  $\bar{P}_n(x)$  the inequality

$$(6) \quad |S_n(x) - \bar{P}_n(x)| > \eta$$

holds. Our aim will be to show that for every  $\eta$  under certain restrictions regarding the given collectives,  $Q_n$  tends toward zero as  $n$  tends toward infinity.

For a fixed point  $x'$  the probabilities  $P_r(x) = p_r$  and  $P_{\mu\nu}(x) = p_{\mu\nu}$  are constants and we put  $\bar{P}_n(x) = \bar{p}_n = (p_1 + p_2 + \dots + p_n)/n$ . The probability that in  $x'$

$$(7) \quad |S_n(x') - \bar{P}_n(x')| > \eta/2$$

is then, according to (4) smaller than  $(s_n^2)_{x'}/(\frac{1}{2}\eta)^2 n^2$ . Here we denote by  $(s_n^2)_{x'}$  the value of  $s_n^2$  in  $x'$  (as given by (1) in paragraph 1).

Now we divide the interval  $(A, B)$  in  $N$  parts in such a way that in every one of the  $N$  intervals e.g. in  $(x', x'')$  the variation

$$(8) \quad \delta = \bar{P}_n(x'') - \bar{P}_n(x') \leq \eta/2.$$

If there is at  $x'$  (or at  $x''$ ) a step of  $\bar{P}_n(x)$  we take the limit which  $\bar{P}_n(x)$  approaches as  $x \rightarrow x'$  (or  $x''$ ) from the interior of the interval. In order to obtain such a division we need only divide the total variation 1 of  $\bar{P}_n(x)$  in  $2/\eta$  equal parts and project these points of division on  $\bar{P}_n(x)$ , disposing however in a suitable way of horizontal parts of  $\bar{P}_n(x)$ . The abscissae of these points form the endpoints of the  $N$  intervals. If there is a step of  $\bar{P}_n(x)$  at an endpoint of one of these intervals the variation in both the adjacent intervals can only be diminished. It is further possible that the two ends of an interval coincide  $x' = x''$ , this will be so if  $\bar{P}_n(x)$  has for  $x'$  a step  $> \eta/2$ . In any case we have a division in  $N \leq 2/\eta$  intervals such that all the points of continuity of  $\bar{P}_n(x)$  are enclosed in them and in each of these intervals (8) is valid.

Let us now assume that in the left end point  $x'$  of the  $r$ th interval  $(x', x'')$  the inequality

$$(9) \quad |S_n(x') - \bar{P}_n(x')| \leq \eta/2$$

is valid. Then we have for every  $x$  between  $x'$  and  $x''$

$$(10) \quad |S_n(x) - \bar{P}_n(x)| \leq \eta/2 + \delta \leq \eta.$$

Because, since  $S_n(x)$  and  $\bar{P}_n(x)$  are both monotone, the difference  $S_n(x') - \bar{P}_n(x')$  cannot increase by more than  $\delta \leq \eta/2$  as  $x$  varies from  $x'$  to  $x''$ . Therefore if (6) is valid for any point  $x$  in this interval then (7) must be valid for the left end point  $x'$  of this interval and the probability  $q_r$  of this latter inequality is less than or equal to  $4(s_n^2)_{x'}/\eta^2 n$ .

But there are  $N$  intervals with the left endpoints  $x'_1, x'_2, \dots, x'_N$  and the

probability that (6) may be valid in any point belonging to any one of these intervals is  $\leq q_1 + q_2 + \dots + q_N$ . Denoting by  $s_n^2$  the greatest of the  $N$  variances  $(s_n^2)_{x_1}, (s_n^2)_{x_2}, \dots, (s_n^2)_{x_N}$  we have for  $Q_n$  (which is the probability that (6) may be valid at any point of continuity of  $P(x)$ ) the inequality

$$(11) \quad Q_n \leq q_1 + q_2 + \dots + q_N \leq \frac{4N}{\eta^2 n^2} s_n^2 \leq \frac{8}{\eta^2} \frac{s_n^2}{n^2}.$$

Therefore  $Q_n$  tends toward zero for every  $\eta$  if  $s_n/n$  tends toward zero.

But according to (2) in paragraph 1,  $s_n/n$  tends toward zero if for every  $x$  in  $(A, B)$

$$(12) \quad \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{\mu, \nu=1}^n [P_{\mu\nu}(x) - P_\mu(x)P_\nu(x)] = 0.$$

Considering the definition of continuity of a statistical function we have obtained the following result:

As in (1'), (2), (3) and (5) let  $P_{\mu\nu}(x, y)$  be two dimensional distributions ( $\mu, \nu = 1, \dots, n; \mu \neq \nu$ ), uniformly bounded in  $(A, B)$ ;  $P_{\mu\nu}(x, B) = P_\mu(x)$ ;  $P_{\mu\nu}(x, x) = P_{\mu\nu}(x)$  and  $\bar{P}_\nu(x) = 1/\nu(P_1(x) + P_2(x) + \dots + P_\nu(x))$ .

If the variable partition  $S_n(x)$  is bounded in  $(A, B)$  and if  $f\{S_n(x)\}$  is uniformly continuous at the "points"  $P_1(x), P_2(x), \dots$  then the probability that

$$(13) \quad |f\{S_n(x)\} - f\{\bar{P}_n(x)\}| > \epsilon$$

tends toward zero for every  $\epsilon$  as  $n \rightarrow \infty$ , provided (12) is uniformly valid for every  $x$  in  $(A, B)$ .

#### 4. Examples. Let us illustrate by simple examples.

1) In order to define the  $P_\nu(x)$  etc. mentioned in our theorem we define the  $n$ -dimensional distribution  $P(x_1, x_2, \dots, x_n)$  used at the beginning of paragraph 3 by indicating the probability density

$$(1) \quad \begin{aligned} \mu(x_1, x_2, \dots, x_n) &= C_n[1 - x_1 x_2 \dots x_n] && \text{in the "unit cube",} \\ &= 0 && \text{elsewhere.} \end{aligned}$$

The corresponding probability distribution is

$$(2) \quad P(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} \mu(x_1, x_2, \dots, x_n) dx_1 \dots dx_n.$$

By putting

$$(3) \quad C_n = \frac{2^n}{2^n - 1},$$

we see that  $P(x_1, x_2, \dots, x_n)$  equals unity if all the arguments are  $\geq 1$  and it equals zero if one of these arguments is less than 0. Therefore  $P(x_1, x_2, \dots, x_n)$  is bounded in the unit cube.

From (1) we deduce the two-dimensional densities

$$(4) \quad \begin{aligned} v_{\mu\nu}(x, y) &= C_n \left(1 - \frac{xy}{2^{n-2}}\right) && \text{in the unit square,} \\ &= 0 && \text{elsewhere} \end{aligned}$$

and the distributions

$$(5) \quad P_{\mu\nu}(x, y) = \int_{-\infty}^x \int_{-\infty}^y v_{\mu\nu}(x, y) dx dy.$$

We see that

$$\begin{aligned} P_{\mu\nu}(x, y) &= C_n xy \left(1 - \frac{xy}{2^n}\right) && \text{in the unit square} \\ &= 0 && \text{if } x \text{ or } y \leq 0 \\ &= 1 && \text{if } x \text{ and } y \geq 1 \end{aligned}$$

and e.g. for  $x \geq 1$ ,  $0 < y < 1$  we have  $P_{\mu\nu}(x, y) = P_{\mu\nu}(1, y)$  etc. Thus the  $P_{\mu\nu}(x, y)$  are completely given.

It follows from (3) that  $-C_n/2^n = 1 - C_n$ ; therefore putting  $C_n = C$  we have in  $(0, 1)$

$$(6) \quad \begin{aligned} P_{\mu\nu}(x, x) &= P_{\mu\nu}(x) = Cx^2 + (1 - C)x^4 \\ P_{\nu}(x) &= Cx + (1 - C)x^2 \end{aligned}$$

therefore

$$(7) \quad P_{\mu\nu}(x) - P_{\mu}(x)P_{\nu}(x) = C(1 - C)x^2(1 - x)^2$$

is  $< 0$  for every  $x$  in  $(0, 1)$  since  $C > 1$ . For  $x \leq 0$ ,  $P_{\mu\nu}(x)$  and  $P_{\nu}(x)$  both equal zero and for  $x \geq 1$  they both equal 1. Therefore our conditions of paragraph 1 are fulfilled. We see that  $C_n$  tends towards unity as  $n \rightarrow \infty$ , therefore for every  $x$  in  $(0, 1)$   $P_{\mu\nu}(x) - P_{\mu}(x)P_{\nu}(x)$  tends towards zero, we have "convergence towards independence" but by no means independence.

This example was based on a *symmetric density*. Let us give an example of asymmetric and *arithmetic* distributions. For the sake of simplicity let  $P_1(x)$ ,  $P_2(x)$ ,  $\dots$  be arithmetic distributions each with only three steps at  $x = 0, 1$  and  $2$ . As starting point we take the  $n$ -dimensional arithmetic distribution  $v(x_1, x_2, \dots, x_n)$  which gives the probability that the first result equals  $x_1$ , the second  $x_2$ ,  $\dots$ , the  $n$ th  $x_n$ , the  $x_i$  being equal to 0 or 1 or 2; thus  $v(x_1, x_2, \dots, x_n)$  takes  $3^n$  values the sum of which equals unity. We deduce the two dimensional distributions  $v_{\mu\nu}(x, y)$ , e.g.  $v_{12}(x, y) = \sum_{x_3, \dots, x_n} v(x, y, x_3, \dots, x_n)$ , the probability that the first result equals  $x$ , the second  $y$ , and finally the  $v_1(x) = \sum v_{12}(x, y)$ , etc. According to the definitions of  $P_{\nu}(x)$  and  $P_{\mu\nu}(x)$  we have then:

$$\begin{aligned}
 (8) \quad P_r(x) &= 0 & (x < 0) \\
 &= v_r(0) & (0 \leq x < 1) \\
 &= v_r(0) + v_r(1) & (1 \leq x < 2) \\
 &= 1 & (2 \leq x), \\
 (9) \quad P_{\mu\nu}(x) &= 0 & (x < 0) \\
 &= v_{\mu\nu}(0, 0) & (0 \leq x < 1) \\
 &= v_{\mu\nu}(00) + v_{\mu\nu}(10) + v_{\mu\nu}(01) + v_{\mu\nu}(11) & (1 \leq x < 2) \\
 &= 1 & (2 \leq x).
 \end{aligned}$$

Now we subject  $v(x_1, \dots, x_n)$  to the following conditions: Every  $v(x_1, \dots, x_n)$  equals zero if it contains either: at least two "zeros," or: at least one "zero" and one "one," or: at least two "ones." All the other  $v$ -values are supposed to be different from zero. Then we have

$$v_{\mu\nu}(0, 0) = v_{\mu\nu}(1, 0) = v_{\mu\nu}(0, 1) = v_{\mu\nu}(1, 1) = 0$$

therefore  $P_{\mu\nu}(x) = 0$  for  $x < 2$  and  $P_{\mu\nu}(x) = 1$  for  $x \geq 2$ . On the other hand  $v_r(0) = v(2, 2, \dots, 2, 0, 2, \dots, 2)$  and  $v_r(1) = v(2, 2, \dots, 2, 1, 2, \dots, 2)$  therefore  $P_r(x) \neq 0$  for  $0 \leq x < 2$  and we have thus for every finite  $n$

$$\begin{aligned}
 P_{\mu\nu}(x) - P_\mu(x)P_\nu(x) &= 0 \quad \text{for } x < 0 \text{ and } x \geq 2, \\
 &< 0 \quad \text{for } 0 \leq x < 2.
 \end{aligned}$$

Therefore the condition (b) of paragraph 1 is fulfilled and thus (12) paragraph 3 holds.

I hope to have the opportunity to discuss more general applications of this theorem later.

A generalization of the *strong* law of large numbers may be given in a similar way.

#### REFERENCES

- [1] B. H. CAMP, *The Mathematical Part of Elementary Statistics*, New York, 1934, page 256.
- [2] R. v. MISES, "Die Gesetze der großen Zahl für statistische Funktionen," *Monatshefte für Math. u. Physik*, 1936, p. 105-128.
- [3] H. GEIRINGER, "Sur les variables aléatoires arbitrairement liées," *Revue de l'Union Interbalcanique*, 1938, p. 6.

BRYN MAWR COLLEGE,  
BRYN MAWR, PENNSYLVANIA.



# CONDITIONS FOR UNIQUENESS IN THE PROBLEM OF MOMENTS

BY M. G. KENDALL

It was shown by Stieltjes [1] that in some circumstances it is possible for two different frequency distributions to have the same set of moments. For instance, the integral

$$\int .4^{n+3} e^{-z} e^{iz} dz$$

around a contour consisting of the positive  $x$ -axis, the infinite quadrant and the positive  $y$ -axis is seen to be zero and it follows that

$$\int_0^{\infty} x^n e^{-x^{\frac{1}{2}}} \sin x^{\frac{1}{2}} dx = 0.$$

Thus the frequency distribution

$$(1) \quad dF = \frac{1}{6} e^{-x^{\frac{1}{2}}} (1 - \lambda \sin x^{\frac{1}{2}}) dx \quad \begin{array}{l} 0 \leq x \leq \infty, \\ 0 \leq \lambda \leq 1 \end{array}$$

has moments which are independent of  $\lambda$ , and equation (1) may be regarded as defining a whole family of distributions each of which has the same moments. It is easy to see that moments of all orders exist, and in fact

$$\mu_r' \text{ (about the origin) } = \frac{1}{6} (4r + 3)!$$

A second example of the same kind, also due to Stieltjes, is the distribution

$$(2) \quad dF = \frac{1}{e^{\frac{1}{2}} \sqrt{\pi}} x^{-\log x} \{1 - \lambda \sin (2\pi \log x)\} dx \quad \begin{array}{l} 0 \leq x \leq \infty, \\ 0 \leq \lambda \leq 1, \end{array}$$

for which

$$\mu_r' = e^{\frac{1}{2}r(r+2)}.$$

The question naturally arises, what are the conditions under which a given set of moments determines a frequency distribution uniquely? The question is of great interest to mathematicians, being closely linked with problems in the theory of asymptotic series, continued fractions and quasi-analytic functions; and it also has importance for statisticians since there is sometimes occasion to be satisfied that a problem of finding a frequency distribution has been uniquely solved by the ascertainment of its moments or semi-invariants. Stieltjes himself considered a more general problem: given a set of constants  $c_0$ ,

$c_1, \dots, c_r, \dots$  does there exist a function  $F$ , non-decreasing and possessing an infinite number of points of increase, such that

$$(3) \quad \int_0^\infty x^r dF =$$

and under what conditions is  $F$  unique, except for an additive constant? Stieltjes showed that if we express the series

$$(4) \quad \sum_{r=0}^{\infty} (-1)^r \frac{c_r}{z^r}$$

as a continued fraction of the form

$$(5) \quad \cfrac{1}{a_1 z +} \cfrac{1}{a_2 +} \cfrac{1}{a_3 z +} \cfrac{1}{a_4 +} \dots \cfrac{1}{a_{2n-1} z +} \cfrac{1}{a_{2n} +} \dots$$

it is a necessary and sufficient condition for the existence of at least one  $F$  that all the  $a$ 's be positive; and that the function is unique or not according as the series  $\sum_{r=0}^{\infty} (a_r)$  diverges or converges. (If the  $a$ 's are positive it must do one or the other.) The integral of equation (3) is to be interpreted in the general Stieltjes sense, so that the result applies to discontinuous as well as to continuous distributions. This is also true of the results obtained below.

Hamburger [2] discussed the similar problem when the limits of the integral in equation (3) are  $\pm \infty$ , and showed that a function  $F$  exists if the expression of (4) as a continued fraction of the form

$$\cfrac{b_0}{a_0 + z +} \cfrac{b_1}{a_1 + z +} \cfrac{b_2}{a_2 + z +} \dots$$

gives positive values of the  $b$ 's. In order that  $F$  may be unique it is necessary and sufficient that the continued fraction be completely (vollständig) convergent in the sense defined by Hamburger.

Unfortunately these criteria, though mathematically complete, are not very useful to statisticians because as a rule it is too difficult to express the coefficients  $a$  and  $b$  explicitly enough in terms of the given  $c$ 's to enable questions of sign or of convergence to be decided. So far as I know, no more convenient criterion for the general Stieltjes problem has been found; but progress is possible if one considers the narrower question: given a set of moments, is the distribution which furnished them unique, that is to say, can any other distribution have furnished them? This is more limited than the Stieltjes problem because we know that at least one solution exists.

Contributions to this subject have been made by Lévy [3] and Carleman [4]. Lévy shows that if moments of all orders exist and are positive it is a sufficient condition for them to determine a distribution uniquely that  $\mu_n^{1/n}/n$  remains finite as  $n$  tends to infinity. (Here and elsewhere in this paper  $\mu_r$  refers to the moment of order  $r$  about any point, not necessarily the mean.) Carleman shows

that, for the case of limits  $-\infty$  to  $+\infty$  the moments determine the distribution uniquely if

$$\sum_{r=0}^{\infty} \frac{1}{(\mu_{2r})^{1/(2r)}}$$

diverges. For the limits 0 to  $\infty$  he gives the corresponding series

$$\sum_{r=0}^{\infty} \frac{1}{(\mu_r)^{1/(2r)}}$$

a criterion which can be improved upon, as will be shown below.

The purpose of this paper is to develop criteria of this kind more systematically and to give more general criteria suitable in cases where the moments are not known explicitly but the behavior of the frequency distribution at its terminals is known.

Three preliminary points necessary for the later argument may be noted.

(1) Define the absolute moment of order  $r$  by

$$\int_{-\infty}^{\infty} |x^r| dF$$

and recall that

$$\nu_1 \leq \nu_2^{1/2} \leq \nu_3^{1/3} \leq \dots \leq \nu_r^{1/r} \leq \dots$$

(cf. Hardy and others, [5]). In other words the quantities  $\nu_r^{1/r}$  form an increasing positive sequence and their reciprocals a decreasing positive sequence.

(2) The quantity  $\nu_n^{1/n}/n$  must either tend to a limit or diverge to infinity as  $n \rightarrow \infty$ . For suppose that

$$\overline{\lim} \nu_n^{1/n}/n = k,$$

$$\underline{\lim} \nu_n^{1/n}/n = l.$$

Writing temporarily  $\nu_n^{1/n} = a_n$ , we have that, given  $\epsilon$  there is an  $N$  such that

$$a_n/n > k - \epsilon$$

for an infinity of values of  $n$  greater than  $N$ . Similarly there is an  $M$  such that

$$a_n/n < l + \epsilon$$

for an infinity of values of  $n$  greater than  $M$ . Now choose  $\rho$  such that  $a_\rho$ ,  $a_{\rho+1}$  are two consecutive values, one near the upper limit and one near the lower limit. This can always be done and we can take  $\rho$  as large as we please. We then have

$$a_\rho > \rho(k - \epsilon)$$

$$a_{\rho+1} < (\rho + 1)(l + \epsilon)$$

and hence, since  $a_{p+1} \geq a_p$

$$(k - \epsilon)\rho < (\rho + 1)(l + \epsilon)$$

giving

$$(k - l) < \frac{l}{\rho} + 2\epsilon + \frac{\epsilon}{\rho}.$$

Thus  $k - l$  can be made as small as we please and is thus zero.

The argument can be very simply adapted to the case in which  $k$  is infinite, and if  $l$  is not finite  $k$ , being not less than  $l$ , is infinite. Thus as  $n \rightarrow \infty$  either  $\lim a_n/n$  exists or  $a_n/n \rightarrow \infty$ .<sup>1</sup>

(3) If any moment fails to converge, so will all moments of higher order. It is evident that more than one distribution can exist having a limited number of finite moments given and the remainder infinite. Thus we need only consider the case when moments of all orders exist. Furthermore, if any even moment exists the absolute moment of next lowest order must exist; for if  $\int_{-\infty}^{\infty} x^{2n} dF$  exists, then each of  $\int_{-\infty}^0 x^{2n} dF$  and  $\int_0^{\infty} x^{2n} dF$  exist separately, each being positive. Hence  $\int_{-\infty}^0 x^{2n-1} dF$  and  $\int_0^{\infty} x^{2n-1} dF$  exist separately and thus  $\int_{-\infty}^{\infty} |x^{2n-1}| dF = -\int_{-\infty}^0 x^{2n-1} dF + \int_0^{\infty} x^{2n-1} dF$  exists. Hence we need only consider the case in which absolute moments of all orders exist.

**THEOREM 1.** *A set of moments determines a distribution uniquely if the series  $\sum_{r=0}^{\infty} \frac{\nu_r t^r}{r!}$  converges for some real non-zero  $t$ .*

Consider the characteristic function

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} dF.$$

This is uniformly continuous in  $t$ , and so are its derivatives of all orders. Thus we have, in the neighborhood of  $t = 0$  the Maclaurin expansion

$$\begin{aligned} \phi(t) &= \sum_{r=0}^{\infty} \left\{ \frac{t^r}{r!} \left[ \frac{d^r \phi}{dt^r} \right]_{t=0} \right\} + R \\ &= \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \mu_r + R. \end{aligned}$$

<sup>1</sup> This proof is necessary to the use of limits in the following theorems, but Theorems 2 and 3 are equally valid if  $\lim$  is substituted for  $\lim$  therein. It is not generally true that if  $a_n$  and  $b_n$  are increasing monotonic sequences either  $\lim a_n/b_n$  exists or  $a_n/b_n \rightarrow \infty$  as  $n \rightarrow \infty$ .

Consequently, under the condition of the theorem, which implies that  $\Sigma \frac{(it)^r}{r!} \mu_r$  is absolutely convergent for some radius  $\rho$ ,  $\phi(t)$  has a Taylor expansion in the neighborhood of the origin and is thus uniquely determined by the moments for  $t < \rho$ . Furthermore, in the neighborhood of  $t = t_0$  we have

$$\phi(t) = \sum_{r=0}^{\infty} \left\{ \frac{i^r (t - t_0)^r}{r!} \int_{-\infty}^{\infty} x^r e^{it_0 x} dF \right\} + R.$$

The modulus of the coefficient of  $\frac{(t - t_0)^r}{r!}$  is not greater than  $\nu_r$ . Therefore  $\phi(t)$  can be expanded in the neighborhood of  $t = t_0$  in a Taylor series with a radius of convergence at least equal to  $\rho$ . Hence the function defining  $\phi(t)$  in the neighborhood of the origin can be continued analytically throughout the range  $-\infty$  to  $+\infty$  and  $\phi(t)$  is uniquely determined in that range.

But the characteristic function uniquely determines the distribution; and hence the theorem follows.

As a result of Theorem 1 we have the following generalization of the criterion given by Lévy.

**THEOREM 2.** *A set of moments completely determines a distribution if  $\lim_{n \rightarrow \infty} \nu_n^{1/n}/n$  is finite.*

It has already been seen that unless  $\nu_n^{1/n}/n$  becomes infinite the limit exists.

By the Cauchy test for convergence the series  $\Sigma \frac{\nu_r t^r}{r!}$  converges if

$$(7) \quad \lim_{n \rightarrow \infty} \left( \frac{\nu_n t^n}{n} \right)^{1/n} < 1.$$

As  $n \rightarrow \infty$ ,  $(n!)^{1/n}$  tends, in accordance with Stirling's theorem, to  $(\sqrt{2\pi n} e^{-n} n^n)^{1/n}$  i.e. to  $n/e$ . Consequently the condition (7) becomes

$$\lim [\nu_n^{1/n}/n] e t < 1.$$

Thus if  $\lim \nu_n^{1/n}/n = k$ , say, the inequality (7) is satisfied for  $t < 1/(ek)$  and the theorem follows.

An important corollary, which enables us to disregard the absolute moments (which may not be given if part of the range is negative) is

**THEOREM 3.** *A set of moments uniquely determines a distribution if  $\lim_{n \rightarrow \infty} \mu_{2n}^{1/(2n)}/n$  is finite.*

$$\text{For} \quad \nu_{2n-1}^{1/(2n-1)} \leq \nu_{2n}^{1/(2n)} = \mu_{2n}^{1/(2n)}.$$

$$\begin{aligned} \text{Thus,} \quad \lim \frac{1}{2n-1} \cdot \nu_{2n-1}^{1/(2n-1)} &\leq \lim \frac{2n}{2n-1} \cdot \frac{1}{2n} \mu_{2n}^{1/(2n)} \\ &\leq \lim \frac{1}{2n} \mu_{2n}^{1/(2n)} \end{aligned}$$

and is therefore finite if the limit on the right is finite. Thus  $\lim \nu_n^{1/n}/n$ , which cannot be greater than the greater of the two limits of  $\nu_{2n-1}^{1/(2n-1)}/(2n-1)$  and  $\nu_{2n}^{1/(2n)}/(2n)$ , must be finite; and the theorem follows from Theorem 2.

Now consider the series  $\sum_{r=0}^{\infty} \frac{1}{\nu_r^{1/r}}$ . Since the successive terms form a monotonic sequence it is a sufficient as well as a necessary condition for convergence that  $n/\nu_n^{1/n}$  tend to zero. Thus, if the series is divergent  $n/\nu_n^{1/n}$  cannot tend to zero and so  $\nu_n^{1/n}/n$  cannot become infinite. Hence it must tend to a finite limit, which may in particular be zero. Hence from Theorem 3 we get

**THEOREM 4.** *A frequency distribution is uniquely determined by its moments if  $\sum_{r=0}^{\infty} \frac{1}{\nu_r^{1/r}}$  diverges.*

Since  $1/\nu_r^{1/r}$  is a decreasing sequence the series  $\sum 1/\nu_r^{1/r}$  converges or diverges with  $\sum 1/\mu_{2r}^{1/(2r)}$ . The Carleman criterion, given by him for the case of limits  $\pm \infty$ , follows. For the case of limits 0 to  $\infty$  the absolute moments are the same as the moments and the criterion can be the divergence of either  $\sum 1/\mu_r^{1/r}$  or  $\sum 1/\mu_{2r}^{1/(2r)}$ . Since  $\mu_r$  is greater than unity in the type of case under consideration the former series provides a more stringent test than that given by Carleman.

At first sight it is rather surprising that the uniqueness of the distribution depends only on the behavior of the even moments, particularly when, by a simple extension of the above result, it is seen that a sufficient condition for uniqueness is the divergence of  $\sum 1/\mu_{4n}^{1/(4n)}$  or  $\sum 1/\mu_{mn}^{1/(mn)}$  or any infinite subset chosen from the moments. It will, however, be remembered that the odd moments are conditioned to some extent by the even moments, and that uniqueness is really determined by the limiting form of  $\nu_n$  as  $n$  tends to infinity.

It is evident that other tests may be derived from Theorem 1 by using the various tests for the convergence of an infinite series. For instance it is a sufficient condition for a set of moments to determine uniquely a distribution with positive range that

$$\frac{\mu_n}{n!} / \frac{\mu_{n+1}}{(n+1)!} = 1 + \frac{\alpha}{n} + O\left(\frac{1}{n^{1+\beta}}\right), \quad \text{where } \begin{matrix} \alpha > 1 \\ \beta > 0 \end{matrix}$$

i.e. that

$$(8) \quad \frac{\mu_n}{\mu_{n+1}} = 1 + \frac{\gamma}{n} + O\left(\frac{1}{n^{1+\beta}}\right), \quad \gamma > 0.$$

It may be noted in passing that the distribution

$$dF = e^{-x} dx \quad 0 \leq x \leq \infty,$$

for which

$$\mu_r \text{ (about origin) } = r!$$

is completely determined by its moments. In fact, by direct reference to Theorem 1 we see that the series  $\sum (it)^r$  converges for  $t < 1$ .

A frequency distribution of finite range is uniquely determined by its moments. For if the range is 0 to  $A$  we have

$$\int_0^A x^r dF \leq A^r$$

and hence  $1/\mu_r^{1/r} \geq 1/A$  so that the series  $\sum 1/\mu_r^{1/r}$  is divergent.

A proof for the case when the frequency distribution is continuous has been given by Lévy, though on entirely different lines from the above.

**THEOREM 5.** *A frequency distribution of infinite range is uniquely determined by its moments if it tends to zero at the infinite terminals faster than  $e^{-x}$ .*

Consider first of all the case when only one end of the range is infinite, so that we may take the range to be 0 to  $\infty$ .

If  $(\mu_n/n!)^{1/n}$  has a finite limit the distribution is unique, by Theorem 2. We have then only to consider the cases (if any) in which  $(\mu_n/n!)^{1/n}$  tends to infinity. It will be shown that in fact such cases do not occur.

Given any (small)  $\epsilon$  there exists an  $X$  such that

$$\frac{f(x)}{e^{-x}} < \epsilon, \quad x > X$$

where  $f(x)$  is the distribution. Thus

$$(9) \quad \int_X^\infty f(x)x^n dx < \epsilon \int_X^\infty e^{-x}x^n dx < \epsilon n!.$$

This is true for all  $n$  and  $X$  is independent of  $n$ . Now,

$$\int_0^\infty f(x)x^n dx = \int_0^X f(x)x^n dx + \int_X^\infty f(x)x^n dx.$$

The first integral on the right is not greater than  $X^n$ . The integral on the left tends, for large  $n$ , to something of greater order than  $n!$ , by our hypothesis, and hence to something of greater order than  $n^n$ . This is of greater order than  $X^n$  (since  $X$ , however large, is independent of  $n$ ) and consequently the second integral on the right is also of greater order than  $n!$ . But this is contrary to equation (9).

The case for the range which is infinite in both directions may be dealt with similarly.

It is easily seen that the two examples of equations (1) and (2) do not tend to infinity faster than  $e^{-x}$ .

Except for the general result of Stieltjes, all the above criteria provide sufficient conditions, but whether the condition of Theorem 1 is also necessary is not certain. An inquiry into the circumstances in which the moment-series of Theorem 1 does not converge throws some light on the question.

It will be remembered that the characteristic function always exists and is uniformly continuous in  $t$ . Since the moments of all orders are assumed to exist we always have

$$\left[ \frac{d^r}{dt^r} \phi(t) \right]_{t=0} = (i)^r \mu_r.$$

Thus, if  $\phi(t)$  can be expanded in an infinite Taylor series that series must be  $\sum \frac{(it)^r}{r!} \mu_r$ . And if this series does not converge then  $\phi(t)$  cannot be expanded as an infinite Taylor series. But it can always be expanded in the finite form with remainder

$$\phi(t) = \sum_{r=0}^r \frac{(it)^r}{r!} \mu_r + R.$$

Thus, when the series does not converge,  $\phi(t)$  can be expanded in powers of  $t$  only asymptotically.

Now it is known that there exist an infinite number of functions which have a given set of coefficients in an asymptotic expansion; for instance, if  $\psi(t)$  has an asymptotic expansion in  $t$  the functions  $\psi(t) + \lambda t^{-\log t}$  all have the same expansion. It is therefore hardly surprising that when the conditions of Theorem 1 break down there can be more than one frequency distribution with the same set of moments.

But it does not follow from what has been said that there *must* be more than one frequency distribution. There must be more than one function, but those functions may not qualify as frequency distributions, e.g. they may be negative in part of the range. In the example just given  $t^{-\log t}$  cannot be a characteristic function, for it does not obey the well-known condition that  $\phi(t)$  and  $\phi(-t)$  should be conjugate.

However, the question is more of mathematical than of statistical interest since the criteria provided above are likely to be adequate for the distributions encountered in practice. For example they establish the uniqueness of the Pearson curves (including the normal curve), the Poisson and the binomial. It would seem that distributions like those of equations (1) and (2) will appear only as statistical curiosities.

#### REFERENCES

- [1] J. STIELTJES, *Recherches sur les fractions continues*, Oeuvres, Groningen, 1918.
- [2] H. HAMBURGER, "Über eine Erweiterung des Stieltjesschen Momentproblems," *Math. Annalen*, Vol. 81 (1920), p. 235, and Vol. 82 (1921), pp. 120 and 168.
- [3] P. LÉVY, *Calcul des Probabilités*, 1925, Paris.
- [4] T. CARLEMAN, *Les Fonctions Quasi-analytiques*, 1925, Paris.
- [5] G. H. HARDY, J. E. LITTLEWOOD and G. PÓLYA, *Inequalities*, 1934, Cambridge, England.

LONDON, ENGLAND.



# ON SAMPLES FROM A NORMAL BIVARIATE POPULATION

By C. T. Hsu

**1. Introduction.** In a number of papers written during the last ten years, J. Neyman and E. S. Pearson<sup>1</sup> have discussed certain general principles underlying the choice of tests of statistical hypotheses. They have suggested that any formal treatment of the subject requires in the first place the specification of (i) the hypothesis to be tested, say  $H_0$ , (ii) the admissible alternative hypotheses. An appropriate test will then consist of a rule to be applied to observational data, for rejecting  $H_0$  in such a way that (iii) the risk of rejecting  $H_0$  when it is true is fixed at some desired value (e.g., 0.05 or 0.01), (iv) the risk of failing to reject  $H_0$  when some one of the admissible alternatives is true is kept as small as possible. With these general principles in mind, they have investigated how best the condition (iv) may be satisfied in different classes of problems. In many cases, though not in all, it has been found that the conditions are satisfied by the test obtained from the use of what has been termed the likelihood ratio, [9], [10], [14]. Once the problem has been specified, the test criterion is usually very easily found, although its sampling distribution, if  $H_0$  is true, often presents great difficulties. In the present paper, I propose to use this method to obtain appropriate tests for a number of hypotheses concerning two normally correlated variables. The investigation was suggested by a recent application of the method by W. A. Morgan [6] to a problem originally discussed by D. J. Finney [3].

**2. The hypotheses and the appropriate criteria.** A sample of two variables  $x_1$  and  $x_2$  is supposed to have been drawn at random from a normal bivariate population, with the distribution

$$(1) \quad p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp \left\{ -\frac{1}{2(1-\rho_{12}^2)} \left[ \left( \frac{x_1 - \xi_1}{\sigma_1} \right)^2 - 2\rho_{12} \left( \frac{x_1 - \xi_1}{\sigma_1} \right) \left( \frac{x_2 - \xi_2}{\sigma_2} \right) + \left( \frac{x_2 - \xi_2}{\sigma_2} \right)^2 \right] \right\}$$

where  $\xi_1$ ,  $\xi_2$ ,  $\sigma_1$ ,  $\sigma_2$ , and  $\rho_{12}$  are the population parameters.

Morgan tested the hypothesis that the variances of the two variables are equal, i.e.,

$$H_1: \quad \sigma_1 = \sigma_2.$$

<sup>1</sup> See bibliography at the end of the paper.

Other hypotheses that will be considered in the present paper are as follows:

- $H_2$  Assuming  $\sigma_1 = \sigma_2$  ; to test  $\rho_{12} = \rho_0$  .
- $H_3$  Assuming  $\sigma_1 = \sigma_2$  ; to test  $\xi_1 = \xi_2$  .
- $H_4$  To test simultaneously  $\sigma_1 = \sigma_2$  ,  $\rho_{12} = \rho_0$  .
- $H_5$  To test simultaneously  $\sigma_1 = \sigma_2$  ,  $\xi_1 = \xi_2$  .
- $H_6$  Assuming  $\sigma_1 = \sigma_2$  and  $\xi_1 = \xi_2$  ; to test  $\rho_{12} = \rho_0$  .
- $H_7$  Assuming  $\sigma_1 = \sigma_2$  , and  $\rho_{12} = \rho_0$  ; to test  $\xi_1 = \xi_2$  .

*Derivation of the criteria.* Let  $x_{1i}$ ,  $x_{2i}$  be the measurements of the two characters on the  $i$ th individual of the sample, then the joint elementary probability law of the two sets of  $n$  observations  $E = (x_{11}, x_{12}, \dots, x_{1n}; x_{21}, x_{22}, \dots, x_{2n})$  is

$$(2) \quad p(E | \xi_1, \xi_2, \sigma_1, \sigma_2, \rho_{12}) = \left( \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \right)^n \cdot \exp \left\{ -\frac{1}{2(1-\rho_{12}^2)} \sum_{i=1}^n \left[ \left( \frac{x_{1i} - \xi_1}{\sigma_1} \right)^2 - 2\rho_{12} \left( \frac{x_{1i} - \xi_1}{\sigma_1} \right) \left( \frac{x_{2i} - \xi_2}{\sigma_2} \right) + \left( \frac{x_{2i} - \xi_2}{\sigma_2} \right)^2 \right] \right\}.$$

It will be convenient to denote by  $A, B, C, D$ , the following conditions of the population from which the sample is supposed to be drawn.

- (A) that stated in equation (1).
- (B) that stated in the equation for  $H_1$ , namely
 
$$\sigma_1 = \sigma_2 = \sigma (\sigma \text{ being unspecified}).$$
- (C)  $\xi_1 = \xi_2 = \xi (\xi \text{ being unspecified}).$
- (D)  $\rho_{12} = \rho_0$  .

Neyman and Pearson's method affords a simple rule for obtaining appropriate test criteria once two sets of conditions have been defined. These are

- (a) the conditions which can be assumed to be satisfied in any case, and
- (b) the conditions which are satisfied if the hypothesis to be tested is true.

The conditions (a) define a class  $\Omega$  of admissible populations, and the conditions (b) define a sub-class  $\omega$  of  $\Omega$  to which the population must belong if the hypothesis tested be true.

The maximum value of  $p(E | \xi_1, \xi_2, \sigma_1, \sigma_2, \rho_{12})$  when the parameters vary in such a way that the population sampled always belongs to  $\Omega$ , is called  $p(\Omega \text{ max.})$ . The maximum value when the population is restricted to  $\omega$  is called  $p(\omega \text{ max.})$ . The likelihood ratio for testing the hypothesis specifying the subset  $\omega$  has been defined to be

$$(3) \quad \lambda = \frac{p(\omega \text{ max.})}{p(\Omega \text{ max.})}.$$

It will be seen that  $1 \leq \lambda \leq 0$ . By referring  $\lambda$ , or a monotonic function of  $\lambda$ , to its sampling distribution when the hypothesis tested is true, we obtain a scale on which to assess our judgment of the truth of the hypothesis tested.

For each of the hypotheses  $H_1$  to  $H_7$ ,  $\lambda$  of (3) can be found. However, we shall use a more convenient criterion.

$$(4) \quad L = \lambda^{2/n}$$

which is a monotonic function of  $\lambda$ .

Thus the respective test criteria are found to be:

For  $H_1$ :

$$(5) \quad L_1 = \frac{4s_1^2 s_2^2 (1 - r_{12}^2)}{(s_1^2 + s_2^2)^2 (1 - R_1^2)}$$

where  $R_1 = \frac{2r_{12}s_1s_2}{s_1^2 + s_2^2}$  is the estimate of  $\rho_{12}$  when  $\sigma_1$  and  $\sigma_2$  are assumed to be equal.

For  $H_2$ :

$$(6) \quad L_2 = \frac{(1 - \rho_0^2)(1 - R_1^2)}{(1 - \rho_0 R_1)^2}.$$

For  $H_3$ :

$$(7) \quad L_3 = 1 / \left\{ 1 + \frac{(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2 - 2r_{12}s_1s_2} \right\}.$$

For  $H_4$ :

$$(8) \quad L_4 = \frac{4(1 - \rho_0^2)s_1^2 s_2^2 (1 - r^2)}{(s_1^2 + s_2^2)^2 (1 - \rho_0 R_1)^2} = L_1 \times L_2.$$

For  $H_5$ :

$$(9) \quad L_5 = \frac{4s_1^2 s_2^2 (1 - r_{12}^2)}{\{s_1^2 + s_2^2 + \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2\} (1 - R_2^2)} = L_1 \times L_3.$$

For  $H_6$ :

$$(10) \quad L_6 = \frac{(1 - \rho_0^2)(1 - R_2^2)}{(1 - \rho_0 R_2)^2}$$

where  $R_2 = \frac{2r_{12}s_1s_2 - \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2 + \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2}$  is the estimate of  $\rho_{12}$  when both the  $\sigma$ 's and the  $\xi$ 's are assumed to be equal.

For  $H_7$ :

$$(11) \quad L_7 = 1 / \left\{ 1 + \frac{(1 + \rho_0)(\bar{x}_1 - \bar{x}_2)^2}{2(s_1^2 - 2\rho_0 r_{12}s_1s_2 + s_2^2)} \right\}^2.$$

The different hypotheses are also given in Table V, at the end of this paper,

together with the conditions defining sets of  $\Omega$  and  $\omega$ , and the appropriate likelihood criteria.

To complete the solution we must find the distributions of  $L$  or some monotonic function of  $L$  in each case when the hypothesis tested is true, in order to assess the significance of an observed value of  $L$ .

**3. The distributions of the criteria.** In order to simplify the problem of finding the distributions of the criteria, consider the following transformation:

$$(12) \quad \begin{aligned} x_{1i} &= (X_i - Y_i)/\sqrt{2} \\ x_{2i} &= (X_i + Y_i)/\sqrt{2}. \end{aligned}$$

It is clear that in view of (1)  $X$  and  $Y$  will be two normally correlated variables. We shall denote this property by  $A'$  corresponding to  $A$ . The conditions  $B'$ ,  $C'$ ,  $D'$  corresponding to  $B$ ,  $C$ ,  $D$  respectively are as follows:

$$\begin{aligned} B': \quad & \rho_{XY} = 0, \\ C': \quad & \xi_X = 0, \\ D': \quad & \sigma_Y^2 = \gamma_0 \sigma_X^2 \quad (\text{when } \rho_{XY} = 0) \end{aligned}$$

where

$$(13) \quad \gamma_0 = \frac{1 + \rho_0}{1 - \rho_0}.$$

Thus we have the equivalent hypotheses  $H'_1, H'_2 \dots H'_7$  corresponding to  $H_1, H_2, \dots H_7$ . The likelihood ratios  $L'_1, L'_2 \dots L'_7$  may be determined in the same way as before, and, in view of the transformation (12), it will be seen that they are equal to  $L_1, L_2 \dots L_7$  respectively.

The tests of the hypotheses  $H'_1, H'_2, H'_3$  are now seen to be well known.

The test of  $H'_1: \rho_{XY} = 0$  is the test for significance of a correlation coefficient, and the criterion  $L_1$  becomes

$$(14) \quad L_1 = \lambda_{H'_1}^{2/n} = 1 - r_{XY}^2.$$

This test has been dealt with by Morgan [6] and Pitman [15], and has been referred to above.

The test of  $H'_2: \sigma_Y^2/\sigma_X^2 = \gamma_0$  when  $\rho_{XY} = 0$  can be treated as an extension of Fisher's  $z$ -test [5], since  $\gamma_0$  is specified. If we write

$$(15) \quad u = \frac{S_Y^2}{S_X^2} = \frac{1 + R_1}{1 - R_1} = \frac{s_1^2 + s_2^2 + 2r_{12}s_1s_2}{s_1^2 + s_2^2 - 2r_{12}s_1s_2}$$

the test criterion  $L_2$  of (6) may be written

$$(16) \quad L_2 = \frac{4u}{\gamma_0(1 + u/\gamma_0)^2}.$$

It is well known that if  $H'_2$  is true, then

$$(17) \quad p(u) = \frac{1}{\gamma_0 B[\frac{1}{2}(n-1), \frac{1}{2}(n-1)]} \left(\frac{u}{\gamma_0}\right)^{\frac{1}{2}(n-2)} \left(1 + \frac{u}{\gamma_0}\right)^{-(n-1)}$$

and the test appropriate to  $H'_2$  and therefore of  $H_2$  is the associated  $z$ -test ( $z = \frac{1}{2} \log u/\gamma_0$ ) with degrees of freedom  $f_1 = f_2 = n - 1$ . It may be easily shown that the two values of  $u$  cutting off equal tail areas from the distribution  $p(u)$  will correspond to a single value of  $L_2$ .

The test of  $H'_3: \xi_X = 0$  when  $\rho_{XY} = 0$  is in the form of "Student's"  $t$  test. If we write

$$(18) \quad \frac{t^2}{n-1} = \frac{\bar{X}^2}{s_X^2} = \frac{(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2 - 2r_{12}s_1s_2}$$

it follows that the test criterion  $L_3$  of (12) may be written

$$(19) \quad L_3 = 1 / \left(1 + \frac{t^2}{n-1}\right).$$

But it is well known that if  $\xi_X = 0$ , then

$$(20) \quad p(t) = \frac{1}{\sqrt{n-1} B[\frac{1}{2}, \frac{1}{2}(n-1)]} \left(1 + \frac{t^2}{n-1}\right)^{-\frac{1}{2}n}.$$

The 5% or 1% points of significance of  $t$  may be obtained from Fisher's  $t$ -table [5] with degrees of freedom  $f = n - 1$ .

The tests of  $H_4$  and  $H_5$ . We infer from (14), (16) and (19) that  $L_1$  is a function of  $r_{XY}$ ,  $L_2$  a function of  $S_Y$  and  $S_X$ , and  $L_3$  a function of  $X$  and  $S_X$ . It is clear that if  $r_{XY}$  is distributed independently of  $S_X$  and  $S_Y$ , then  $L_1$  and  $L_2$  are independent, i.e.,

$$(21) \quad p(L_1, L_2) = p(L_1)p(L_2)$$

and that if  $r_{XY}$  is distributed independently of  $X$  and  $S_X$ , then  $L_1$  and  $L_3$  are independent, i.e.,

$$(22) \quad p(L_1, L_3) = p(L_1)p(L_3).$$

It is known that  $X, Y$  are independent of  $S_X, S_Y, r_{XY}$ ; and in addition that  $r_{XY}$  is distributed independently of  $S_X, S_Y$  if  $\rho_{XY} = 0$ . Therefore, if  $H'_1$  is true, then the relations (21) and (22) hold. Hence, knowing  $p(L_1)$  and  $p(L_2)$ , a very simple transformation and integration gives  $p(L_4)$ . Similarly, the distribution of  $L_5$  may be readily derived from those of  $L_1$  and  $L_3$ .

But from the distribution of  $r_{XY}$  when  $\rho_{XY} = 0$ , by transformation (14), the distribution of  $L_1$  assuming  $H'_1$  true is found to be

$$(23) \quad p(L_1) = \frac{1}{B[\frac{1}{2}(n-2), \frac{1}{2}]} L_1^{\frac{1}{2}(n-4)} (1 - L_1)^{-\frac{1}{2}}.$$

If  $H'_2$  is true, from (17), by transformation (16) we have

$$(24) \quad p(L_2) = \frac{1}{B[\frac{1}{2}(n-1), \frac{1}{2}]} L_2^{\frac{1}{2}(n-1)} (1-L_2)^{-\frac{1}{2}}$$

Again, if  $H'_2$  is true, from (20), by transformation (19), we have

$$(25) \quad p(L_2) = \frac{1}{B[\frac{1}{2}(n-1), \frac{1}{2}]} L_2^{\frac{1}{2}(n-1)} (1-L_2)^{-\frac{1}{2}}$$

which is the same as the distribution of  $L_2$ . Therefore by comparing (21) and (22) we see that the distribution of  $L_5$  when  $H'_5$  is true will be exactly the same as that of  $L_4$  when  $H'_4$  is true. We shall therefore confine ourselves to the problem of obtaining the distribution of  $L_4$  from those of  $L_1$  and  $L_2$ .

Now

$$(26) \quad p(L_1, L_2) = \frac{1}{B[\frac{1}{2}(n-2), \frac{1}{2}]B[\frac{1}{2}(n-1), \frac{1}{2}]} L_1^{\frac{1}{2}(n-2)} (1-L_1)^{-\frac{1}{2}} L_2^{\frac{1}{2}(n-1)} (1-L_2)^{-\frac{1}{2}}$$

Applying the transformation

$$(27) \quad \begin{aligned} L_4 &= L_1 L_2 \\ Z &= L_2 \end{aligned}$$

and integrating with respect to  $Z$  from 0 to 1, we obtain

$$(28) \quad p(L_4) = \frac{1}{2}(n-2)L_4^{\frac{1}{2}(n-4)}, \quad 0 \leq L_4 \leq 1.$$

Thus we can construct the values of  $L_4$  at the 5% and 1% levels for different values of  $n$  as given in Table I.

TABLE I  
5% and 1% values of  $L_4$  (or  $L_5$ )

$n$	5%	1%
5	.1357	.0464
6	.2509	.1000
7	.3017	.1585
8	.3684	.2154
9	.4249	.2683
10	.4729	.3162
12	.5493	.3981
15	.6307	.4924
20	.7169	.5995
24	.7616	.6579
30	.8074	.7197
40	.8541	.7848
60	.9019	.8532
120	.9505	.9249
$\infty$	1.0000	1.0000

*The test of  $H_6$ .* In the case of testing  $H'_6(\sigma_Y^2 = \gamma_0 \sigma_X^2)$ , assuming  $\rho_{XY}$  and  $\rho_X$  each to be zero, the likelihood estimate of  $\sigma_X^2$  becomes  $\Sigma X^2/n$  or  $S_X^2 + \bar{X}^2$ . The distribution of this quantity is the same as that of  $S_X^2$  but with degrees of freedom  $n$  instead of  $n - 1$ . Therefore, by analogy with the previous result (17) used in testing  $H_2$ , if we write

$$(29) \quad v = \frac{nS_Y^2}{\Sigma X^2} = \frac{S_Y^2}{S_X^2 + \bar{X}^2} = \frac{1 + R_2}{1 - R_2}$$

then the likelihood criterion of  $H_6$  becomes

$$(30) \quad L_6 = \frac{4v}{\gamma_0 \left(1 + \frac{v}{\gamma_0}\right)^2}$$

and

$$(31) \quad p\left(v \mid \frac{\sigma_Y^2}{\sigma_X^2} = \gamma_0\right) = \frac{1}{\gamma_0 B\left[\frac{1}{2}(n-1), \frac{1}{2}n\right]} \left(\frac{v}{\gamma_0}\right)^{\frac{1}{2}(n-1)} \left(1 + \frac{v}{\gamma_0}\right)^{-(n-1)}.$$

Hence the test appropriate to  $H_6$  is the associated  $z$ -test  $z = \frac{1}{2} \log \left\{ \frac{v}{\gamma_0} \middle/ \frac{n-1}{n} \right\}$  with  $f_1 = n - 1$ ,  $f_2 = n$ . We can use the  $z$ -table as before.

*The test of  $H_7$ .* Here we test whether  $\xi_X = 0$ . It may be seen that  $L_7$  is a function of  $\bar{X}^2/(S_Y^2 + \gamma_0 S_X^2)$ . Further, if we assume that  $\rho_{XY} = 0$  and also that  $\sigma_Y^2 = \gamma_0 S_X^2$ , then it will follow that  $\Sigma(X - \bar{X})^2$  and  $\frac{1}{\gamma_0} \Sigma(Y - \bar{Y})^2$  are each distributed independently as  $\chi^2 \sigma_X^2$  with  $n - 1$  degrees of freedom; and hence their sum is distributed as  $\chi^2 \sigma_X^2$  with  $2n - 2$  degrees of freedom. Also if  $\xi_X = 0$  (and  $H'_7$  is true)  $X$  will be distributed normally about zero with standard error  $\sigma_X/\sqrt{n}$ . Hence we may write

$$(32) \quad L_7 = 1 \middle/ \left\{ 1 + \frac{t^2}{2n-2} \right\}^2$$

where

$$(33) \quad t^2 = \bar{X}^2 \middle/ \sqrt{\frac{\Sigma(X - \bar{X})^2 + \Sigma(Y - \bar{Y})^2/\gamma_0}{n(2n-2)}}$$

and is distributed in accordance with "Student's" distribution with  $2n - 2$  degrees of freedom,

$$(34) \quad p(t_2) = \frac{1}{\sqrt{2n-2} B\left[\frac{1}{2}, \frac{1}{2}(2n-2)\right]} \left(1 + \frac{t^2}{2n-2}\right)^{-\frac{1}{2}(2n-1)}.$$

In terms of original variables

$$(35) \quad \frac{t_2^2}{2n-2} = \frac{\gamma_0 \bar{X}^2}{\gamma_0 S_X^2 + S_Y^2} = \frac{(1 + \rho_0)(\bar{x}_1 - \bar{x}_2)^2}{2(s_1^2 - 2\rho_0 r_{12} s_1 s_2 + s_2^2)}.$$

4. Comparison of the  $R_1$ -test and  $R_2$ -test with the  $r_{12}$ -test in cases where  $H_1$  and  $H_0$  are true respectively. It will be noted that in the preceding discussion we have been concerned with three different tests of the hypothesis that  $\rho_{12}$  has some specified value  $\rho_0$ . When there is no information available regarding the means and standard deviations of  $x_1$  and  $x_2$ , the test is based on the sampling distribution of the ordinary product-moment coefficient  $r_{12}$ . If it may be assumed that  $\sigma_1 = \sigma_2$ , then we have the estimate

$$R_1 = \frac{2r_{12}s_1s_2}{s_1^2 + s_2^2}.$$

If besides  $\sigma_1 = \sigma_2$ , it may also be assumed that  $\xi_1 = \xi_2$ , then we have the estimate

$$R_2 = \frac{2r_{12}s_1s_2 - \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2 + \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2}.$$

From the point of view of testing hypotheses, all these criteria  $r_{12}$ ,  $R_1$ ,  $R_2$  follow from the application of the likelihood ratio method. It will be noted that if  $\sigma_1 = \sigma_2$ , either the  $r_{12}$  or the  $R_1$  test may be used. But, insofar as the likelihood principle is accepted, the latter should be regarded as the "better" test. Again, if  $\sigma_1 = \sigma_2$  and  $\xi_1 = \xi_2$ , all three tests may be used, but that based on  $R_2$  will be the "best". A question of interest is to investigate just what is meant by the "better" or the "best" test. We may ask how far the improvements are sufficient to justify the use of the  $R_1$  and  $R_2$  tests in place of the more generally used  $r_{12}$  test. One method of comparison is to examine what Neyman and Pearson [12] have termed the "power function" of the tests.

For example, when testing the hypothesis that a parameter  $\theta$  has the value  $\theta_0$  in the population sampled, the power of the test criterion  $T$  with regard to the alternative hypothesis that  $\theta = \theta_1 > \theta_0$  is given by the expression  $\beta(\theta_1) = P\{T > T_\alpha | \theta = \theta_1\}$  where  $T_\alpha$  is the value of  $T$  at the level of significance  $\alpha$ . This quantity  $\beta(\theta)$  measures the chance that the test as specified will detect the fact that  $\theta = \theta_0$ , i.e., the chance of rejecting the hypothesis when it is not true. A test whose power function is never less than that of any other test is termed the uniformly most powerful test.

If the permissible alternative hypotheses to  $\theta = \theta_0$  are both  $\theta < \theta_0$  and  $\theta > \theta_0$ , then the power of the test  $T$  is given by the expression

$$\beta(\theta_1) = 1 - p\{T'_\alpha < T < T''_\alpha | \theta_1\}$$

where  $T'_\alpha$  and  $T''_\alpha$  are the values of  $T$  at both ends of the distribution at the level of the significance  $\alpha$ . When the test is such that the power function has a minimum value  $\alpha$  at  $\theta = \theta_0$ , it is said to be unbiased.

A test is termed biased if, for certain alternative hypotheses  $\theta \neq \theta_0$ , the chance of rejecting the hypothesis  $\theta = \theta_0$  is less than the chance of rejecting this hypothesis when it is true.



In what follows it is proposed to compare the power functions of the tests based on  $r_{12}$ ,  $R_1$ , and  $R_2$  in order to obtain more complete evidence of the extent to which one is "better" than the other.

*The distribution of  $R_1$ .*<sup>2</sup> We have obtained the distribution of  $n$  when  $H'_2$  and therefore  $H_2$  is true. We are now able to find the distribution of  $R_1$  by applying the transformation of (15). Thus the distribution of  $R_1$  in terms of  $\rho_0$  is

$$(36) \quad p(R_1|\rho_0) = \frac{(1 - \rho_0^2)}{2^{n-2} B[\frac{1}{2}(n-1), \frac{1}{2}(n-1)]} \frac{(1 - R_1^2)^{\frac{1}{2}(n-2)}}{(1 - \rho_0 R_1)^{n-1}}.$$

The significance of  $R_1$  may be assessed by the  $z$ -test, where we take

$$(37) \quad Z - \frac{1}{2} \log_e \frac{u}{\gamma_0} = \frac{1}{2} \log \frac{1 + R_1}{1 - R_1} - \frac{1}{2} \log \frac{1 + \rho_0}{1 - \rho_0} \\ = z' - \zeta, \text{ say}$$

with degrees of freedom  $f_1 = f_2 = n - 1$ . R. A. Fisher's  $z$ -table may be used in this connection.

When  $\rho_{12} = 0$ , the distribution simplifies to

$$(38) \quad p(R_1|\rho_{12} = 0) = \frac{1}{2^{n-1} B[\frac{1}{2}(n-1), \frac{1}{2}(n-1)]} (1 - R_1^2)^{\frac{1}{2}(n-2)} \\ = \frac{1}{B[\frac{1}{2}(n-1), \frac{1}{2}]} (1 - R_1^2)^{\frac{1}{2}(n-2)}$$

since  $2^{2^{n-2}} B[\frac{1}{2}(n-1), \frac{1}{2}(n-1)]$  is equal to  $B[\frac{1}{2}(n-1), \frac{1}{2}]$  by duplication formula [16, p. 240].

The distribution (38) is similar in form to that of  $p(r_{12}|\rho_{12} = 0)$  with  $n - 1$  degrees of freedom instead of  $n - 2$ . The significance levels of  $R_1$  may then be obtained directly from the  $r$ -table [1] for the case  $\rho_{12} = 0$ , entering with degrees of freedom  $n - 1$ .

*The distribution of  $R_2$ .* The distribution of  $R_2$  may be obtained from that of  $v$  when  $H'_6$  and therefore  $H_6$  is true. It is

$$(39) \quad p(R_2|\rho_{12} = \rho_0) = \frac{(1 + \rho_0)^{\frac{1}{2}n} (1 - \rho_0)^{\frac{1}{2}(n-1)}}{2^{n-1} B[\frac{1}{2}(n-1), \frac{1}{2}n]} \frac{(1 + R_2)^{\frac{1}{2}(n-2)} (1 - R_2)^{\frac{1}{2}(n-2)}}{(1 - \rho_0 R_2)^{n-1}}.$$

This agrees with the result first obtained by R. A. Fisher [4] by a different method. The significance of  $R_2$  may be assessed by the  $z$ -test, where we take

$$(40) \quad z = \frac{1}{2} \log \left( \frac{v}{\gamma_0} / \frac{n-1}{n} \right)$$

<sup>2</sup> Since finding the distribution of  $R_1$  (36), (38) and the relation between  $R_1$  and  $z'$  (37), my attention has been drawn to a recent paper by DeLury [2] in which the same results are obtained. Since my method of derivation is different from his, I have thought it worthwhile to retain it here.

with degrees of freedom  $f_1 = n - 1$ ,  $f_2 = n$ . The tables for use with the  $z$ -test may be used in this connection.

When  $\rho_{12} = 0$ , the distribution is simplified to

$$(41) \quad p(R_2 | \rho_{12} = 0) = \frac{1}{2^{n-1} B[\frac{1}{2}(n-1), \frac{1}{2}n]} (1 + R_2)^{\frac{1}{2}(n-1)} (1 - R_2)^{\frac{1}{2}(n-1)}$$

which is simply a Pearson Type I curve.

*Power functions of  $R_1$  and  $R_2$ .* In order to find the power functions of  $R_1$  and  $R_2$  with respect to alternative hypotheses  $H_1$  to  $H_2$ , specifying  $\rho_{12} = \rho_1 < \rho_0$ , it will be convenient to consider the incomplete beta function distributions

$$(42) \quad p(x_1) = \frac{1}{B[\frac{1}{2}(n-1), \frac{1}{2}(n-1)]} x_1^{\frac{1}{2}(n-1)} (1 - x_1)^{\frac{1}{2}(n-1)}$$

$$(43) \quad p(x_2) = \frac{1}{B[\frac{1}{2}(n-1), \frac{1}{2}n]} x_2^{\frac{1}{2}(n-1)} (1 - x_2)^{\frac{1}{2}(n-1)}$$

where  $x_1 = \frac{u}{\gamma_0(1 + u/\gamma_0)}$  and  $x_2 = \frac{v}{\gamma_0(1 + v/\gamma_0)}$ . From the *Tables of the Incomplete Beta Function* [13] we can find the values of  $x_1$  and  $x_2$  at the significance level  $\alpha$ , i.e.

$$(44) \quad I_{x_1}[\frac{1}{2}(n-1), \frac{1}{2}(n-1)] = \alpha',$$

$$(45) \quad I_{x_2}[\frac{1}{2}(n-1), \frac{1}{2}n] = \alpha'.$$

The values of  $R'_1(\alpha)$ , and of  $R'_2(\alpha)$ , may then be calculated from the relations

$$(46) \quad R_1 = \frac{u-1}{u+1} = \frac{-1+x_1+\gamma_0 x_1}{1-x_1+\gamma_0 x_1},$$

$$(47) \quad R_2 = \frac{v-1}{v+1} = \frac{-1+x_2+\gamma_0 x_2}{1-x_2+\gamma_0 x_2}.$$

The power functions of  $R_1$  and  $R_2$  thus found may be given as follows:

$$(48) \quad \beta'(\rho_1 | R_1) = P\{R_1 < R'_1(\alpha) | \rho_1\},$$

$$(49) \quad \beta'(\rho_1 | R_2) = P\{R_2 < R'_2(\alpha) | \rho_1\}.$$

In the same way, for any alternative hypothesis  $H_1$  specifying  $\rho_{12} = \rho_1 > \rho_0$ , we can find the values of  $x_1$  and  $x_2$  at the significance level  $\alpha''$ , at the other end of the distribution, i.e.

$$(50) \quad 1 - I_{x_1}[\frac{1}{2}(n-1), \frac{1}{2}(n-1)] = \alpha'',$$

$$(51) \quad 1 - I_{x_2}[\frac{1}{2}(n-1), \frac{1}{2}n] = \alpha''.$$

Thence the corresponding values of  $R''_1(\alpha)$  and  $R''_2(\alpha)$  may be obtained, and their power functions are

$$(52) \quad \beta''(\rho_1 | R_1) = P\{R_1 > R''_1(\alpha) | \rho_1\},$$

$$(53) \quad \beta''(\rho_t | R_2) = P\{R_2 > R_2''(\alpha) | \rho_t\}.$$

The power functions of  $R_1$  and  $R_2$  with respect to alternative hypotheses specifying  $\rho_{12} = \rho_t < \rho_0$  and  $> \rho_0$  may now be obtained by adding (48) and (52) or (49) and (53) or, more simply,

$$(54) \quad \beta(\rho_t | R_1) = 1 - P\{R_1'(\alpha) < R_1 < R_1''(\alpha) | \rho_t\},$$

$$(55) \quad \beta(\rho_t | R_2) = 1 - P\{R_2'(\alpha) < R_2 < R_2''(\alpha) | \rho_t\}$$

where  $R_1'(\alpha)$ ,  $R_1''(\alpha)$ ;  $R_2'(\alpha)$ ,  $R_2''(\alpha)$  are the values of  $R_1$  and  $R_2$  at the two ends of the distribution at the significance level  $\alpha = \alpha' + \alpha''$ .

In view of the fact that after transformation the tests based on  $R_1$  and  $R_2$  are equivalent to tests regarding the equality of variances, it follows from Neyman and Pearson's work [11] regarding the uniformly most powerful test of the hypothesis that  $\sigma_Y^2/\sigma_X^2 = \gamma_0$ , with alternatives  $\sigma_Y^2/\sigma_X^2 = \gamma_t < \gamma_0$  (or  $\gamma_t > \gamma_0$ ), that: (1) if  $\sigma_1 = \sigma_2$  and alternative to  $\rho_{12} = \sigma_0$  are that  $\rho_{12} = \rho_t < \rho_0$  (or, in a second case,  $\rho_t > \rho_0$ ) the test based on  $R_1$  is the uniformly most powerful test, i.e., it is more powerful than that based on  $r_{12}$ ; and (2) if  $\sigma_1 = \sigma_2$  and  $\xi_1 = \xi_2$ , then the test based on  $R_2$  is the uniformly most powerful test, i.e., it is more powerful than those based on either  $r_{12}$  or  $R_1$ .

For illustration, let us take a special case, say

$$(a) \quad n = 10, \quad \rho_0 = 0.6, \quad \alpha' = \alpha'' = 0.025.$$

From the tables, we obtain the values

$$\begin{array}{ll} x_1' = .198902 & x_2' = .184863 \\ x_1'' = .801098 & x_2'' = .772916 \end{array}$$

and by calculation the values

$$\begin{array}{ll} R_1'(\alpha) = -.0034 & R_2'(\alpha) = -.0487 \\ R_1''(\alpha) = .8831 & R_2''(\alpha) = .8632. \end{array}$$

The values of the power functions of  $R_1$  and  $R_2$  for specified values of  $\rho_t$  have been calculated and are given in Table II. For  $\rho_t < \rho_0$ , a comparison of columns 2 and 4 will show that the test based on  $R_2$  is uniformly more powerful than that based on  $R_1$  (or for  $\rho_t > \rho_0$ , a comparison of columns 3 and 5).

*The unbiased test of  $H_2$  and  $H_6$ .* When however the alternatives are that  $\rho_{12} = \rho_t < \rho_0$ , and  $\rho_t > \rho_0$ , questions of bias may be introduced.

In the case of  $H_2$ , i.e. when  $R_1$  is used, it was established by J. Neyman in his lecture courses [8], that if we test whether  $\sigma_Y^2/\sigma_X^2 = \gamma_0$ , where the alternatives are  $\gamma_t < \gamma_0$  and  $\gamma_t > \gamma_0$ , and if the samples of  $X$  and  $Y$  are of equal size, then the test based on cutting off equal tail areas of the distribution of  $x_1$  is unbiased and of the type  $B$  [7]. Therefore the same may be said of the  $R_1$ -test.

In the case of  $H_6$ , the equivalent transformed test is again whether  $\sigma_Y^2/\sigma_X^2 = \gamma_0$ . But the test now corresponds to that in which an estimate of  $\sigma_Y^2$  is based

on  $f_1 = n - 1$  degrees of freedom and an estimate of  $\sigma_x^2$  on  $f_2 = n$  degrees of freedom. The degrees of freedom not being equal, it is known that if equal tail areas are cut off from the sampling distribution of  $x_2$ , this test will be biased. Neyman's result [8] shows that if the lower and upper significance levels are taken at  $x'_2$  and  $x''_2$ , then the equation

$$(56) \quad x_2''^{f_1}(1 - x_2'')^{f_2} = x_2'^{f_1}(1 - x_2')^{f_2}$$

should be satisfied if the test is unbiased. Since in the present case, with the test based on equal tail area critical region, the bias will be very small, the rejection levels  $R'_2(\alpha)$  and  $R''_2(\alpha)$  in the numerical investigation given in Table III have been selected taking equal tail areas for simplicity.

TABLE II

*Values of the power functions of  $R_1$  and  $R_2$  with respect to alternative hypotheses*

$$\rho_{12} = \rho_1 < \rho_0 \text{ OR } \rho_1 > \rho_0$$

$$(n = 10; \rho_0 = 0.6; \alpha' = \alpha'' = 0.025)$$

$\rho_1$	$\beta'(\rho_1 R_1)$	$\beta''(\rho_1 R_1)$	$\beta'(\rho_1 R_2)$	$\beta''(\rho_1 R_2)$
-0.8	.9984			
-0.6	.9739		.9807	
-0.4	.9867		.9005	
-0.2	.7189		.7360	
0.0	.4960	.0002	.5093	.0001
0.2	.2744	.0008	.2809	.0006
0.3	.1825	.0018	.1860	.0015
0.4	.1106	.0042	.1111	.0037
0.5	.0576	.0099	.0580	.0093
0.6	.025	.025	.025	.025
0.7	.0081	.0678	.0080	.0720
0.8	.0015	.1995	.0015	.2150
0.9	.0001	.5950	.0001	.6289
0.95		.8979		.9150
0.975		.9866		.9897

If we now take a special case, similar to (a) above, but taking equal tail areas, so that

$$\begin{aligned} n &= 10 & \rho &= 0.6 \\ \alpha &= 0.5 & (\alpha' = \alpha'' = \tfrac{1}{2}\alpha) \end{aligned}$$

we can obtain the values of  $x$ 's and of  $R$ 's as before.

The values of the power functions of  $R_1$  and  $R_2$  for specified values of  $\rho_1$  are given in columns 3 and 4 of Table III. These values are equivalent to the sums of the corresponding values in Table II. The values of the power functions of  $R_1$  and  $R_2$  for the following additional cases are also given in Table III:

$$(b) \quad n = 10 \quad \rho_0 = 0.8 \quad \alpha = 0.05$$

$$(c) \quad n = 20 \quad \rho_0 = 0.6 \quad \alpha = 0.05$$

$$(d) \quad n = 20 \quad \rho_0 = 0.8 \quad \alpha = 0.05.$$

*Comparison of the power functions.* We may now deal with the question raised at the beginning of this section, namely, as to what is meant by the "better" or "best" test. We shall proceed to compare for certain special cases the power functions of the three test, all of which are applicable where it may be assumed that  $\sigma_1 = \sigma_2$ ,  $\xi_1 = \xi_2$ .

In the first place it will be noted that the power function of the test based on equal tail areas of the  $r_{12}$  distribution is

$$(57) \quad \beta(\rho_t | r_{12}) = 1 - p\{\gamma'_{12}(\alpha) < r_{12} < \gamma''_{12}(\alpha) | \rho_t\}$$

where

$$(58) \quad \begin{aligned} P\{r_{12} < \gamma'_{12}(\alpha) | \rho_0\} &= \int_{-1}^{\gamma'_{12}(\alpha)} p(r_{12} | \rho_{12} = \rho_0) dr_{12} = \frac{1}{2}\alpha \\ P\{r_{12} > \gamma''_{12}(\alpha) | \rho_0\} &= \int_{\gamma''_{12}(\alpha)}^1 p(r_{12} | \rho_{12} = \rho_0) dr_{12} = \frac{1}{2}\alpha \end{aligned}$$

and

$$(59) \quad p(r_{12} | \rho_{12} = \rho_0) = \frac{(1 - \rho_0^2)^{\frac{1}{2}(n-1)}}{\pi \Gamma[\frac{1}{2}(n-1)]} (1 - r_{12}^2)^{\frac{1}{2}(n-4)} \left( \frac{\partial}{\partial r_{12}} \right)^{n-2} \frac{\cos^{-1}(-\rho_0 r_{12})}{\sqrt{(1 - \rho_0^2 r_{12}^2)}}$$

The probability that  $r_{12}$  is less than some specified value may be obtained from *Tables of the Correlation Coefficient* (F. N. David, [1]), or, where these are not sufficiently detailed, by using R. A. Fisher's  $z'$ -transformation for  $r_{12}$  [4].

The cases considered are (a), (b), (c), (d) as defined above. The power functions of the three different tests (all based upon the equal tail areas of their distributions) are given in Table III. The figures for  $r_{12}$  in the brackets are those obtained by the  $z'$ -transformation approximation.

An examination of Tables II and III brings out the following points:

(1) For reasons given above, the  $R_2$  test based on equal tail area critical regions is very slightly biased; the amount of this bias for the case  $n = 10$ ,  $\rho_0 = 0.6$ ,  $\alpha = 0.05$  is shown in Table IV. This shows that the power of the  $R_2$  test is less than 0.05 in the fifth or sixth decimal places for  $0.59 < \rho_t < 0.60$ . As a result this test is very slightly less powerful than the other two tests for alternatives with  $\rho_t$  slightly less than  $\rho_0$ . The effect is, however, of little importance.

(2) Except in this short range of  $\rho_t$ , we find that

$$\beta(\rho_t | R_2) \geq \beta(\rho_t | R_1) \geq \beta(\rho_t | r_{12}).$$

TABLE III  
*Comparison of the power functions of  $r_{12}$ ,  $R_1$ , and  $R_2$  tests with respect to alternative hypotheses*

$\rho_i$	$n = 10 \quad \rho_0 = 0.6$			$n = 10 \quad \rho_0 = 0.8$			$n = 20 \quad \rho_0 = 0.6$			$n = 20 \quad \rho_0 = 0.8$		
	$\beta(\rho_i r_{12})$	$\beta(\rho_i R_1)$	$\beta(\rho_i R_2)$	$\beta(\rho_i r_{12})$	$\beta(\rho_i R_1)$	$\beta(\rho_i R_2)$	$\beta(\rho_i r_{12})$	$\beta(\rho_i R_1)$	$\beta(\rho_i R_2)$	$\beta(\rho_i r_{12})$	$\beta(\rho_i R_1)$	$\beta(\rho_i R_2)$
-0.6	.9739	.9739	.9807	.9887	.9891	.9921	.9965	.9967	.9973			
-0.4	.8865	.8867	.9005	.9557	.9569	.9650	.9648	.9663	.9698			
-0.2	.7186	.7189	.7360	.9742	.8766	.8909	.8328	.8369	.8449	.9952	.9959	.9966
0.0	.4960	.4962	.5094	.7158	.7189	.7360	.5412	.5456	.5534	.9624	.9663	.9698
0.2	.2753	.2752	.2815	.4727	.4750	.4877	.2026	.2036	.2061	.8062	.8170	.8254
0.4	.1142	.1148	.1148	.3330	.3345	.3427	.0915	.0917	.0922	.6309	.6432	.6520
0.5	.0679	.0675	.0673	.2005	.2010	.2047	.0500	.0500	.0500	.3920	.4011	.4085
0.6	.0500	.0500	.0500	.0969	.0965	.0971	.1096	.1119	.1147	.1589	.1617	.1635
0.7	.0735	.0759	.0800	.0500	.0500	.0500	.3886	.4010	.4134	.0500	.0500	.0500
0.8	.1890	.2010	.2165	.1466	.1771	.1904	.9034	.9106	.9181	.3272	.3493	.3604
0.9	.5656 (.5569)	.5951	.6290	(.1454)			(.9004)			(.3270)		
0.95	(.8709)	.8979	.9150	(.4689)	.5426	.5763	(.9974)	.9974	.9978	(.8547)	.5871	.8844
0.975	(.9822)	.9866	.9897	(.8134)	.8692	.8896				(.9944)	.9947	.9960
0.99				(.9845)	.9908	.9938						
Levels	-.0039 .8998	-.0034 .8831	-.0487 .8632	.4004 .9574	.3817 .9463	.3423 .9368	.2289 .8300	.2253 .8201	.2041 .8084	.5671 .9222	.5613 .9158	.5459 .9101

That is to say, the power function of the  $R_2$  test never lies below those of the  $R_1$  and  $r_{12}$  tests, and that of the  $R_1$  test never lies below that of the  $r_{12}$  test.

(3) The gain in sensitivity as measured by the chance that the test will detect that  $\rho_1 \neq \rho_0$  is, however, very small. Further,  $R_1$  may only be used if it is known that  $\sigma_1 = \sigma_2$  and  $R_2$  if it is known in addition that  $\xi_1 = \xi_2$ . It will only be in rather special problems that the statistician can feel confident that such assumptions are justified. We will therefore probably prefer the test based on the ordinary product moment correlation coefficient  $r_{12}$ , since the slight loss in power will be felt to be outweighed by the gain in simplicity. It is, however, only after an objective comparison of the consequences of applying the three tests that a definite opinion on these points can be reached.

TABLE IV

$\rho_1$	$\beta'(\rho_1 R_2)$	$\beta''(\rho_1 R_2)$	$\beta(\rho_1 R_2)$
0.5	.0580	.0093	.0673
0.590	.0274235	.0225806	.0500041
0.591	.0271778	.0228190	.0499968
0.592	.0269359	.0230578	.0499937
0.593	.0266934	.0232976	.0499910
0.594	.0264515	.0235337	.0499852
0.595	.0262096	.0237798	.0499894
0.596	.0259677	.0240222	.0499899
0.597	.0257257	.0242651	.0499908
0.598	.0254838	.0245107	.0499945
0.599	.0252419	.0247540	.0499959
0.6	.025	.025	.05

**5. Summary.** Various hypotheses relating to a population of two normal correlated variates have been considered and the appropriate test criteria for each hypothesis have been derived by the likelihood ratio method. The distributions of the likelihood ratio criteria or of monotonic functions of them have been obtained with the aid of transformation (14). References have been given to tables from which significance levels for use in conjunction with the tests may be obtained; a new table of significance levels for the tests of  $H_1$  and  $H_2$  was given.

The power functions of  $r_{12}$ ,  $R_1$  and  $R_2$  have been compared; from these power functions it was concluded that  $R_1$  and  $R_2$  are suitable respectively for testing the hypothesis when  $\sigma_1 = \sigma_2$  and when, in addition,  $\xi_1 = \xi_2$ .

In conclusion, I should like to express my indebtedness to Professor E. S. Pearson for continued advice and help in the preparation of this paper, to Dr. A. Wald and Professor S. S. Wilks for valuable suggestions.

TABLE V  
Conditions defining  $\Omega$  and  $\omega$  together with the likelihood criteria appropriate for testing the hypotheses  $H_i$

(1) Hypotheses $H_i$	(2) Initial Assumptions (Apart from Normality)	(3) To be tested	(4) Conditions defining (a) $\Omega$	(5) Conditions defining (b) $\omega$	(6) Criteria $L_i = \lambda_{H_i}^{2/n}$
$H_1$	None	$\sigma_1 = \sigma_2$	A	A, B	$\frac{4s_1^2 s_2^2 (1 - r_{12}^2)}{\{(s_1^2 + s_2^2)^2 - 4r_{12}^2 s_1^2 s_2^2\}}$
$H_2$	$\sigma_1 = \sigma_2$	$\rho_{12} = \rho_0$	A, B	A, B, D	$\frac{(1 - \rho_0^2)(1 - R_1^2)^1}{(1 - \rho_0 R_1)^2}$
$H_3$	$\sigma_1 = \sigma_2$	$\xi_1 = \xi_2$	A, B	A, B, C	$1 / \left\{ 1 + \frac{(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2 - 2r_{12}s_1s_2} \right\}$
$H_4$	None	$\sigma_1 = \sigma_2$ $\rho_{12} = \rho_0$	A	A, B, D	$\frac{4s_1^2 s_2^2 (1 - \rho_0^2)(1 - r_{12}^2)}{(s_1^2 + s_2^2)(1 - \rho_0 R_1)^2}$
$H_5$	None	$\sigma_1 = \sigma_2$ $\xi_1 = \xi_2$	A	A, B, C	$\frac{4s_1^2 s_2^2 (1 - r_{12}^2)}{\{s_1^2 + s_2^2 + \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2\}(1 - R_2^2)^2}$
$H_6$	$\sigma_1 = \sigma_2$ $\xi_1 = \xi_2$	$\rho_{12} = \rho_0$	A, B, C	A, B, C, D	$\frac{(1 - \rho_0^2)(1 - R_2^2)}{(1 - \rho_0 R_2)^2}$
$H_7$	$\sigma_1 = \sigma_2$ $\rho_{12} = \rho_0$	$\xi_1 = \xi_2$	A, B, D	A, B, D, C	$1 / \left\{ 1 + \frac{(1 + \rho_0)(\bar{x}_1 - \bar{x}_2)^2}{2(s_1^2 - 2\rho_0 r_{12}s_1s_2 + s_2^2)} \right\}^2$

$$^1 R_1 = \frac{2r_{12}s_1s_2}{s_1^2 + s_2^2} \quad ^2 R_2 = \frac{2r_{12}s_1s_2 - \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2 + \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2}$$



## REFERENCES

- [1] F. N. DAVID, *Tables of the Correlation Coefficient*, London: Biometrika Office, 1938.
- [2] D. B. DELURY, *Ann. Math. Stat.* Vol. 9 (1938) p. 149.
- [3] D. J. FINNEY, *Biometrika*, Vol. 30 (1938), p. 190.
- [4] R. A. FISHER, *Metron*, Vol. 1 (1921), p. 3.
- [5] R. A. FISHER, *Statistical Methods for Research Workers*, 7th ed. Edinburgh: Oliver Boyd, 1938.
- [6] W. A. MORGAN, *Biometrika*, Vol. 31 (1939), p. 13.
- [7] J. NEYMAN, *Bull. Soc. Math. France*, Vol. 63 (1935), p. 246.
- [8] J. NEYMAN, Lectures delivered in London, 1937-8, (Unpublished).
- [9] J. NEYMAN AND E. S. PEARSON, *Biometrika*, Vol. 20A (1928), p. 175.
- [10] J. NEYMAN AND E. S. PEARSON, *Bull. Acad. Polonaise Sci. Lettres*, A (1931), p. 460.
- [11] J. NEYMAN AND E. S. PEARSON, *Phil. Trans.*, Ser. A, Vol. 236 (1933), p. 289.
- [12] J. NEYMAN AND E. S. PEARSON, *Stat. Res. Mem.* Vol. 1 (1936), p. 1.
- [13] K. PEARSON (Editor), *Tables of the Incomplete Beta-Function*, London: Biometrika Office.
- [14] E. S. PEARSON AND J. NEYMAN, *Bull. Acad. Polonaise Sci. Lettres*, A (1930), p. 73.
- [15] J. C. PITMAN, *Biometrika*, Vol. 31 (1939), p. 9.
- [16] E. T. WHITTAKER AND G. N. WATSON, *Modern Analysis*, 4th Edition (1927).

UNIVERSITY COLLEGE,  
LONDON, ENGLAND.

## ON A LEAST SQUARES ADJUSTMENT OF A SAMPLED FREQUENCY TABLE WHEN THE EXPECTED MARGINAL TOTALS ARE KNOWN

BY W. EDWARDS DEMING AND FREDERICK F. STEPHAN

**1. Introduction.** There are situations in sampling wherein the data furnished by the sample must be adjusted for consistency with data obtained from other sources or with deductions from established theory. For example, in the 1940 census of population a problem of adjustment arises from the fact that although there will be a complete count of certain characters for the individuals in the population, considerations of efficiency will limit to a sample many of the cross-tabulations (joint distributions) of these characters. The tabulations of the sample will be used to estimate the results that would have been obtained from cross-tabulations of the entire population.<sup>1</sup> The situation is shown in Fig. 1 in parallel tables for the universe and for the sample. For the universe the marginal totals  $N_{i.}$  and  $N_{.j}$  are known, but not the cell frequencies  $N_{ij}$ ; for the sample, however, tabulation gives both the cell frequencies  $n_{ij}$  and the marginal totals  $n_{i.}$  and  $n_{.j}$ .

In estimating any cell frequency of the universe, such as  $N_{ij}$ , three possibilities present themselves; from the sample one may make an estimate from the  $i$ th row alone, another from the  $j$ th column alone, and still another from the over-all ratio  $n_{ij}/n$ : specifically, the three estimates would be  $n_{ij}N_{i.}/n_{i.}$ ,  $n_{ij}N_{.j}/n_{.j}$ , and  $n_{ij}N/n$ . As a result of sampling errors these will not be identical except by accident, and though any of them by itself may be considered accurate enough, still, if the whole  $r \times s$  table of universe cell frequencies were so estimated, the marginal totals would not come out right. In this paper we present a rapid method of adjustment, which in effect combines all three of the estimates just mentioned, and at the same time enforces agreement with the marginal totals. The method is extended to varying degrees of cross-tabulation in three dimensions.

In any problem of adjustment where the conditions are intricate it is necessary to have a method that is straight-forward and self-checking; this becomes imperative when we realize that in the three-dimensional Case VII of the problem now at hand (*vide infra*), any adjustment in one cell must be balanced by adjustments in at least seven others. The method of least squares is one possible procedure for effecting an adjustment and at the same time enforcing certain conditions among the marginal totals. It is essentially a scheme for

<sup>1</sup> Examples will occur in the 1940 census publications. Further discussion of this problem and of the sampling procedure is given by the authors in "The sampling procedure of the 1940 population census," *Jour. Am. Stat. Assn.*, Vol. 35 (1940), pp. 615-630.

arriving at a set of calculated or adjusted observations that will satisfy the conditions of the problem, and at the same time minimize the sum of the weighted squares of the residuals, symbolized as

$$(1) \quad S = \sum w(n_c - n_0)^2$$

$n_c$  and  $n_0$  being the calculated and observed numbers in a cell, and  $n_c - n_0$  the corresponding residual. It is the nature of the conditions imposed on the adjusted values that distinguishes one type of problem from another. Least squares has the practical advantage of uniqueness, once the weights of the observations have been assigned, and it possesses the theoretical dignity of giving one kind of "best" estimates under ideal conditions of sampling. For our present purpose we shall minimize sums of the form

$$(2) \quad S = \sum (m_i - n_i)^2 / n_i$$

$n_i$  being the observed frequency in the  $i$ th cell, and  $m_i$  the calculated or adjusted frequency therein. The conditions among the  $m_i$  will arise from the fact that the marginal totals, after adjustment, must agree with their expected values, namely, the deflated marginal totals of the universe (for example,  $m_{i.}$  and  $m_{.j}$  as defined in eqs. (6) and (7)).

By definition, weight and variance are inversely proportional, hence the principle of least squares is identical with the minimizing of chi-square. Here the variance in the  $i$ th cell is  $\nu_i(1 - \nu_i/n)$ , where  $\nu_i$  is the expected number in that cell, and  $n$  is the total number in the sample. Now if  $\nu_i$  is sufficiently well approximated by  $n_i$ , it follows that if no cell contains an appreciable fraction of the whole sample (a circumstance requiring a fair sized number of cells—perhaps 100), the variance may be taken as  $\nu_i$  for every  $i$ , and the minimized  $S$  can be used as chi-square. But regardless of the number of cells, if the  $n_i$  be not too much different from one another, so that the factor  $1 - \nu_i/n$  may be treated as a constant, we still get the least squares solution by minimizing  $S$  as defined in eq. (2).

**2. The two dimensional problem.** Suppose that the data on two characteristics (e.g. age and highest grade of school completed) are obtained for each member of a universe of  $N$  individuals, and that tabulations of the data provide either (a) one set of marginal totals  $N_{1.}, N_{2.}, \dots, N_{r.}$ ; or (b) in addition, the marginal totals  $N_{.1}, N_{.2}, \dots, N_{.s}$ . The nature of the tabulations is presumed such that it is not feasible to count the numbers  $N_{ij}$  in the cells, as would be done if one character were crossed with the other. Suppose, however, that for a sample of  $n$  individuals selected in a random manner from the universe, the two characters are crossed with each other, so that we know not only all the  $s + r$  marginal totals  $n_{.1}, \dots, n_{r.}$  of the sample but also the numbers  $n_{ij}$  ( $i = 1, 2, \dots, r; j = 1, 2, \dots, s$ ). The problem is to estimate the unknown frequencies  $N_{ij}$  in the cells of the universe. This will be done by finding the calculated or adjusted sample frequencies  $m_{ij}$  and then inflating them by the inverse sampling ratio  $N/n$ .



$$(8) \quad N_{.1} + N_{.2} + \dots + N_{.s} = N_{.1} + N_{.2} + \dots + N_{.r} = N.$$

It is for this reason that if  $i$  runs through all  $r$  values in eq. (4), then  $j$  can run through only  $s - 1$  in eq. (5). A similar equation also exists for the marginal totals of the sample, namely,

$$(9) \quad n_{.1} + n_{.2} + \dots + n_{.s} = n_{.1} + n_{.2} + \dots + n_{.r} = n.$$

*Solution of the two dimensional Case I.* Assuming that the adjusted values of the  $m_{ij}$  have been found, let each take on a small variation  $\delta m_{ij}$ ; then the differentials of eqs. (3) and (4) show that

$$(10) \quad \frac{1}{2}\delta S = \sum \{(m_{ij} - n_{ij})/n_{ij}\}\delta m_{ij} = 0 \quad (\text{one equation}),$$

$$(11) \quad \sum_j \delta m_{ij} = 0, \quad i = 1, 2, \dots, r \quad (r \text{ equations}).$$

Multiply now eq. (11) by the arbitrary Lagrange multiplier  $-\lambda_i$ , and add eqs. (10) and (11) to obtain

$$(12) \quad \sum \{(m_{ij} - n_{ij})/n_{ij} - \lambda_i\}\delta m_{ij} = 0. \quad (\text{one equation}).$$

By the usual argument, one may now set each brace equal to zero, recognizing that the  $r$  Lagrange multipliers are then no longer arbitrary but must satisfy the relation

$$(13) \quad m_{ij} = n_{ij}(1 + \lambda_i).$$

The adjusted frequencies  $m_{ij}$  can be computed at once as soon as the  $\lambda_i$  are found. To evaluate them one may rewrite the conditions (4) using the right-hand member of (13) for  $m_{ij}$ , obtaining

$$(14) \quad m_{i.} = n_{i.}(1 + \lambda_i).$$

Another way to arrive at this same relation is to sum each member of eq. (13) in the  $i$ th row. However obtained  $\lambda_i$  is now known, since  $m_{i.}$  and  $n_{i.}$  are known, and in fact eq. (13) now gives

$$(15) \quad m_{ij} = n_{ij}(m_{i.}/n_{i.}).$$

The adjustment is thus a simple proportionate one by rows, the cells in any one row all being raised or lowered by the proportionate adjustment in the row total. Case I thus amounts to  $r$  independent one dimensional proportionate adjustments, one for each row, and any one or all may be carried out, as desired. This result can be obtained by a simpler approach but is presented in this way for consistency with later cases.

The minimized sum of squares may be computed directly, or from the row totals by seeing that

$$(16) \quad S = \sum_i (m_{i.} - n_{i.})^2/n_{i.}.$$

The term  $(m_{i.} - n_{i.})^2/n_{i.}$  for the  $i$ th row may be considered separately, and

used as  $\chi^2$  with  $s - 1$  degrees of freedom, or all rows may be combined into the minimized  $S$  as given in eq. (16), and used as  $\chi^2$  with  $r(s - 1)$  degrees of freedom.

*Solution of the two dimensional Case II.* In addition to eqs. (11) we now have also

$$(17) \quad \sum_i \delta m_{ij} = 0 \quad j = 1, 2, \dots, s - 1$$

which comes by differentiating eqs. (5). By addition of eqs. (10), (11), and (17), after multiplying eq. (11*i*) by  $-\lambda_i$  and eq. (17*j*) by  $-\lambda_j$ , we obtain

$$(18) \quad \Sigma \{ (m_{ij} - n_{ij}) / n_{ij} - \lambda_i - \lambda_j \} \delta m_{ij} = 0.$$

Equating each brace to zero, as before, we find that

$$(19) \quad m_{ij} = n_{ij}(1 + \lambda_i + \lambda_j)$$

wherein  $\lambda_s$  is to be counted 0. The adjustment is now no longer proportionate by rows, but involves every cell.

To evaluate the Lagrange multipliers in eq. (19) we may sum the two members downward and across in Fig. 1 and obtain the  $r + s - 1$  normal equations

$$(20) \quad \begin{aligned} n_{i.} \lambda_i + \sum_j n_{ij} \lambda_j &= m_{i.} - n_{i.}, \quad i = 1, 2, \dots, r \\ \sum_i n_{ij} \lambda_i + n_{.j} \lambda_j &= m_{.j} - n_{.j}, \quad j = 1, 2, \dots, s - 1. \end{aligned}$$

These can be reduced for numerical computation. The top row solved for  $\lambda_i$  gives

$$(21) \quad \lambda_i = (1/n_{i.}) \{ m_{i.} - \sum_j n_{ij} \lambda_j \} - 1$$

whereupon by substitution into the bottom row of eqs. (20) we arrive at the  $s - 1$  normal equations

$$(22) \quad \begin{array}{ccccccc} \lambda_{.1} & \lambda_{.2} & \dots & \lambda_{.s-1} & & = & 1 \\ \hline n_{.1} - \sum_i \frac{n_{i1} n_{i1}}{n_{i.}} & - \sum_i \frac{n_{i1} n_{i2}}{n_{i.}} & \dots & - \sum_i \frac{n_{i1} n_{i,s-1}}{n_{i.}} & & = & m_{.1} - \sum_i \frac{n_{i1} m_{i.}}{n_{i.}} \\ & n_{.2} - \sum_i \frac{n_{i2} n_{i2}}{n_{i.}} & \dots & - \sum_i \frac{n_{i2} n_{i,s-1}}{n_{i.}} & & = & m_{.2} - \sum_i \frac{n_{i2} m_{i.}}{n_{i.}} \\ & & & \vdots & & & \vdots \\ & & & n_{.s-1} - \sum_i \frac{n_{i,s-1} n_{i,s-1}}{n_{i.}} & = & m_{.s-1} - \sum_i \frac{n_{i,s-1} m_{i.}}{n_{i.}} \\ & & & & & & 0. \end{array}$$

Because of symmetry in the coefficients, those below the diagonal are not shown, indeed, in a systematic computation, they are not used. The 0 in the bottom

row is appended for the computation of the minimized  $S$ , if desired. The number of Lagrange multipliers to be solved for directly is  $s - 1$ , and the remaining ones come by substitution into eq. (21),  $\lambda_s$  being counted 0.

A simple procedure for calculating the coefficients in the normal equations (22) is to set up a preparatory table by dividing each  $n_{ij}$  in the  $i$ th row by  $\sqrt{n_{i.}}$ ; also to write down  $m_{i.}/\sqrt{n_{i.}}$  for that row, for use on the right-hand side of the normal equations (compare Tables I and II). In machine calculation the constant divisor  $\sqrt{n_{i.}}$  would be left on the keyboard until the entire  $i$ th row is divided; or, if reciprocal multiplication is preferred, the multiplier  $1/\sqrt{n_{i.}}$  would be left on. From this preparatory table, the cumulation of squares and cross-products in the vertical gives the required summations for the coefficients. The sum check would be applied in the usual manner.

**3. A numerical example of the two dimensional Case II.** The fact is that in practice one need not bother about forming and solving the normal equations because they will be displaced by a simplifying iterative procedure, to be explained in a later section. For illustration, however, we may do an example both ways, first using the normal equations and the adjustment (19), later on accomplishing the same results by the quicker method.

We may start with the unitalicized numbers in the  $4 \times 6$  array of Table I, assuming these to be the sampling frequencies  $n_{ij}$  to be adjusted. Actually, they were obtained by deflating 1/20th (for a supposed 5 per cent sample) the New England age  $\times$  state table on p. 1108 of vol. 2 of the *Fifteenth Census of the U. S.*, 1930, then varying the deflated values by chance with Tippett's numbers to get our sampling frequencies  $n_{ij}$ . The italicized entries in Table I represent the final (adjusted)  $m_{ij}$ , and it is these that we now set out to get. We start off with the sampling frequencies  $n_{ij}$  and the known marginal totals  $m_{.1}$ ,  $m_{.2}$ , etc., where  $m_{i.} = N_{i.}n/N$ ,  $m_{.j} = N_{.j}n/N$ , as in eqs. (6) and (7). The Lagrange multipliers shown along the left-hand and top borders arise in the calculations now to be undertaken.

Table II is the preparatory table, advised at the close of the last section. It is derived from Table I by dividing the  $i$ th row of sample frequencies by  $\sqrt{n_{i.}}$ . For example, the entry 8.64 in the cell  $i = 3, j = 2$  comes by dividing 419 by  $\sqrt{2352}$ , 419 being the entry in the cell of the same indices in Table I, and 2352 being the sum of the third row. The sums at the bottom and right-hand side are for checking the formation of the normal equations. The cumulations of squares and cross-products along the vertical give the summations required for the normal eqs. (22), which now appear numerically as eqs. (23).

	No.	$\lambda_1$	$\lambda_2$	$\lambda_3$	=	1
	1	7413	-3549	-2354	=	$3197 \times 10^{-4}$
(23)	2		4441	-544	=	2356
	3			3129	=	-3222
	4					0

Performing the solution by any favorite procedure one will obtain

$$(24) \quad \lambda_1 = .01182 \quad \lambda_2 = .01490 \quad \lambda_3 = .00119$$

TABLE I

*A table of artificial sample frequencies, an artificial 5 percent sample of native white persons of native white parentage attending school, by age by state, New England, 1930. The adjusted frequency  $m_{ij}$  in each cell is shown italicized just below the corresponding sample frequency  $n_{ij}$*

Age			7 to 13	14 & 15	16 & 17	18 to 20	
$j =$ $\lambda_j =$			1 .0118	2 .0149	3 .0012	4 0	$n_{.j}$ $m_{.j}$
State	$i$	$\lambda_i$					
Maine	1	-.0146	3623 <i>3613</i>	781 <i>781</i>	557 <i>550</i>	313 <i>308</i>	5274 <i>5252</i>
New Hampshire	2	-.0003	1570 <i>1588</i>	395 <i>401</i>	251 <i>251</i>	155 <i>155</i>	2371 <i>2395</i>
Vermont	3	.0234	1553 <i>1608</i>	419 <i>435</i>	264 <i>270</i>	116 <i>119</i>	2352 <i>2432</i>
Massachusetts	4	-.0162	10538 <i>10492</i>	2455 <i>2452</i>	1706 <i>1680</i>	1160 <i>1141</i>	15859 <i>15766</i>
Rhode Island	5	-.0230	1681 <i>1662</i>	353 <i>350</i>	171 <i>167</i>	154 <i>150</i>	2359 <i>2330</i>
Connecticut	6	-.0034	3882 <i>3915</i>	857 <i>867</i>	544 <i>543</i>	339 <i>338</i>	5622 <i>5662</i>
$n_{.j}$			22847	5260	3493	2237	33837
$m_{.j}$			<i>22877</i>	<i>5285</i>	<i>3462</i>	<i>2213</i>	<i>33837</i>

The adjusted  $m_{ij}$  (italicized) are rounded off, hence when summed may occasionally disagree a unit or so with the expected marginal totals (also italicized), the latter arise by deflation from the universe rather than by direct addition of the  $m_{ij}$ .

whereupon by substitution into eq. (21) comes

$$(25) \quad \begin{aligned} \lambda_1 &= -.0146 & \lambda_4 &= -.0162 \\ \lambda_2 &= -.0003 & \lambda_5 &= -.0230 \\ \lambda_3 &= +.0234 & \lambda_6 &= -.0034. \end{aligned}$$

The next step is to compute the  $m_{ij}$  by eq. (19). Table I is now bordered with the Lagrange multipliers for a convenient arrangement of the factors required, and the calculation is completed. It will be noted that, for example



$$(26) \quad m_{32} = 419(1 + .0234 + .0149) = 435.$$

The  $m_{ij}$  thus calculated are shown italicized in Table I. The marginal totals, found by adding the  $m_{ij}$  just calculated, do not agree exactly everywhere with the expected totals, because of rounding off to integers: the errors of closure, however, are slight, and it is a simple matter to raise or lower some of the larger cells by a unit or two to force exact satisfaction of the conditions, if this is desired.

**4. The three dimensional problem.** Here the  $N$  cards of the universe are sorted and counted for one and perhaps a second and third characteristic, and possibly crossed by pairs in various combinations (Cases I-VII). The sample of  $n$ , however, is crossed by all three characteristics, which is to say that the

TABLE II

*This comes by dividing each sample frequency in Table I by the corresponding  $\sqrt{n_i}$ .  
(This operation would ordinarily be done a row at a time)*

	$j =$				$m_{i.}/\sqrt{n_i}$	Sum
	1	2	3	4		
$i = 1$	49.89	10.75	7.67	4.31	72.32	144.94
2	32.24	8.11	5.15	3.18	49.19	97.87
3	32.02	8.64	5.44	2.39	50.15	98.64
4	83.68	19.49	13.55	9.21	125.19	251.12
5	34.61	7.27	3.52	3.17	47.97	96.54
6	51.77	11.43	7.26	4.52	75.51	150.49
Sum	284.21	65.69	42.59	26.78	420.33	839.60

cell frequencies  $n_{ijk}$  are all known (refer to Fig. 2). As before, the adjusted frequencies are required.

*Case I: One set of slice totals known.* Assume the slice totals  $N_{1..}$ ,  $N_{2..}$ ,  $\dots$ ,  $N_{r..}$  to be known; the conditions are then

$$(27) \quad \sum_{jk} m_{ijk} = m_{i.} = N_{i..} n/N \quad i = 1, 2, \dots, r$$

being  $r$  in number. The summation to be minimized is

$$(28) \quad S = \sum (m_{ijk} - n_{ijk})^2 / n_{ijk}$$

being similar to that in eq. (3), except that now there are three indices to be summed over instead of two. Following a procedure similar to that used before, we differentiate eqs. (27) and (28) and introduce the  $r$  Lagrange multipliers  $\lambda_i$ .

with eq. (27). The steps are identical with those of the two dimensional Case I, and the result is at once

$$(29) \quad m_{ijk} = n_{ijk}(1 + \lambda_{i..}) = n_{ijk}(m_{i..}/n_{i..}).$$

This adjustment, like that shown by eq. (15), is a simple proportionate one, but this time by slices rather than by columns. All cell frequencies having the same  $i$  index are raised or lowered in the same proportion.

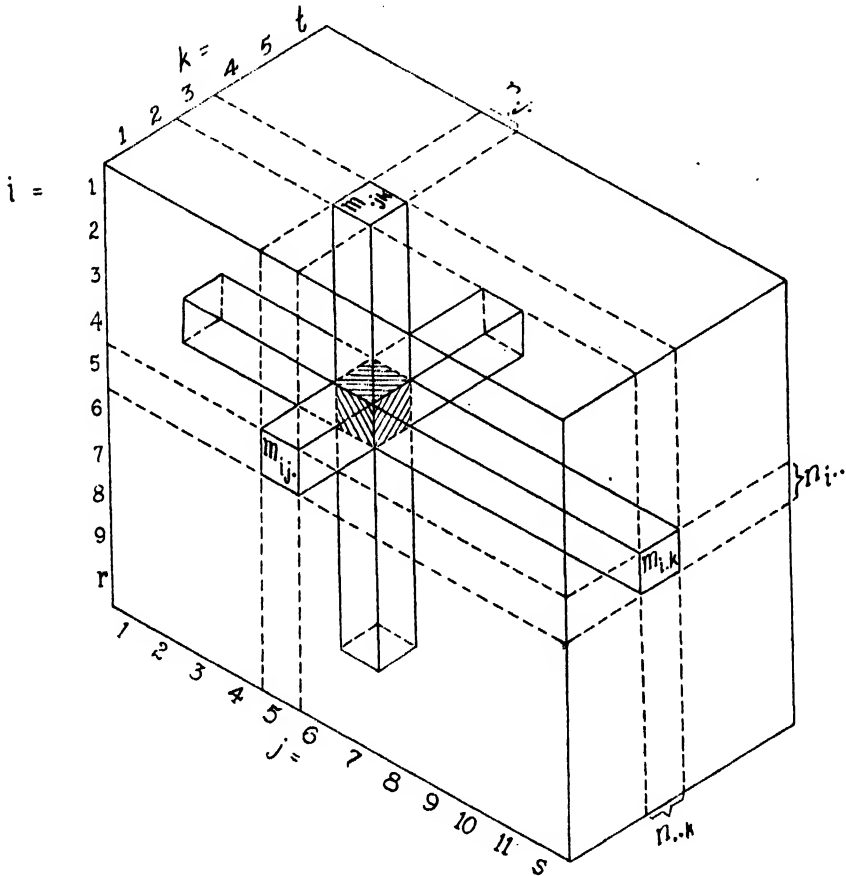


FIG. 2. SHOWING THE SYSTEM OF NOTATION FOR THE CELL FREQUENCIES AND MARGINAL TOTALS IN THE THREE DIMENSIONAL SAMPLE

*Case II: Two sets of slice totals known.* Here, in addition to the slice totals of Case I we know also

$$N_{.1}, N_{.2}, \dots, N_{.s}.$$

whence arise the  $s - 1$  additional conditions

$$(30) \quad \sum_{ik} m_{ijk} = m_{.j.} = N_{.j.} n/N, \quad j = 1, 2, \dots, s - 1.$$

Using the Lagrange multiplier  $\lambda_{.j}$  here, and  $\lambda_{i..}$  with eq. (27) as before, we find that

$$(31) \quad m_{ijk} = n_{ijk}(1 + \lambda_{i..} + \lambda_{.j})$$

in which  $\lambda_{..}$  is to be counted zero. This adjustment is proportionate by tubes, the ratio  $m_{ijk}/n_{ijk}$  being constant along the  $ij$ th tube and in fact equal to  $m_{ij.}/n_{ij.}$ , independent of  $k$ . Unfortunately we do not here know the face totals  $m_{ij.}$  and are unable to make use of the proportionality as we shall in Case IV.

To solve for the  $r + s - 1$  Lagrange multipliers we sum the members of eq. (31) over  $j$  and then over  $i$  and arrive at the normal equations

$$(32) \quad \begin{aligned} n_{i..}\lambda_{i..} + \sum_j n_{ij.}\lambda_{.j} &= m_{i..} - n_{i..}, \quad i = 1, 2, \dots, r, \\ \sum_i n_{ij.}\lambda_{i..} + n_{.j.}\lambda_{.j} &= m_{.j.} - n_{.j.}, \quad j = 1, 2, \dots, s - 1. \end{aligned}$$

These can be reduced to  $s - 1$  equations in precisely the same way that eqs. (20) were reduced, but because of the iterative process to come further on, we shall not pursue the reduction here.

*Case III: All three sets of slice totals known.* All slice totals

$$\begin{aligned} N_{.1.}, N_{.2.}, \dots, N_{.s.} \\ N_{1..}, N_{2..}, \dots, N_{r..} \\ N_{..1}, N_{..2}, \dots, N_{..t} \end{aligned}$$

now being known, in addition to conditions (27) and (30) we require here

$$(33) \quad \sum_{ij} m_{ijk} = m_{..k} = N_{..k}n/N, \quad k = 1, 2, \dots, t - 1$$

which makes a total of  $r + (s - 1) + (t - 1)$  or  $r + s + t - 2$  conditions. The same kind of manipulation as used heretofore gives

$$(34) \quad m_{ijk} = n_{ijk}(1 + \lambda_{i..} + \lambda_{.j} + \lambda_{..k})$$

with  $\lambda_{..}$  and  $\lambda_{..t}$  to be counted zero. The adjustment is no longer proportionate by slices or tubes, but involves every cell. In practice, once the normal equations are solved and the Lagrange multipliers worked out, one proceeds very much as in the two dimensional Case II: for each of the  $t$  slices, corresponding to the  $t$  values of  $k$ , there will be a two dimensional adjustment, the 1 in eq. (19) being replaced now by  $1 + \lambda_{..k}$ .

The normal equations for the Lagrange multipliers can be found by performing double summations on eq. (34). The result is

$$(35) \quad \begin{aligned} n_{i..}\lambda_{i..} + \sum_j n_{ij.}\lambda_{.j} + \sum_k n_{i.k}\lambda_{..k} &= m_{i..} - n_{i..}, \quad i = 1, 2, \dots, r, \\ \sum_i n_{ij.}\lambda_{i..} + n_{.j.}\lambda_{.j} + \sum_k n_{.jk}\lambda_{..k} &= m_{.j.} - n_{.j.}, \quad j = 1, 2, \dots, s - 1, \\ \sum_i n_{i.k}\lambda_{i..} + \sum_j n_{.jk}\lambda_{.j} + n_{..k}\lambda_{..k} &= m_{..k} - n_{..k}, \quad k = 1, 2, \dots, t - 1. \end{aligned}$$

If these calculations were to be carried out, one would simplify the computation by solving the top row for  $\lambda_{i..}$ , getting

$$(36) \quad \lambda_{i..} = (1/n_{i..}) \{ m_{i..} - \sum_j n_{ij} \lambda_{.j} - \sum_k n_{i.k} \lambda_{..k} \} - 1$$

and then substituting this into the middle and last rows of eqs. (35) to get a reduced set of  $s + t - 2$  normal equations for the Lagrange multipliers  $\lambda_{.j}$  and  $\lambda_{..k}$ , the numerical values of which when set back into eq. (36) give the  $\lambda_{i..}$ . In all the summations of eqs. (35) and (36),  $\lambda_{..}$  and  $\lambda_{..i}$  would be counted zero. But here again, the iterative process to be explained later will displace the use of normal equations, so actually we are not interested in reducing them.

*Case IV: One set of face totals known.* It may be that the  $rs$  face totals

$$N_{11.}, N_{12.}, \dots, N_{ij.}, \dots, N_{rs.}$$

are known from crossing the  $i$  and  $j$  characters in the universe. The conditions are then

$$(37) \quad \sum_k m_{ijk} = m_{ij.} = N_{ij.} n/N \quad \begin{array}{l} i = 1, 2, \dots, r, \\ j = 1, 2, \dots, s. \end{array}$$

The adjustment here turns out to be

$$(38) \quad m_{ijk} = n_{ijk}(1 + \lambda_{ij.});$$

but by summing both sides over the index  $k$  to evaluate  $\lambda_{ij.}$  it is seen that

$$(39) \quad m_{ij.} = n_{ij.}(1 + \lambda_{ij.}),$$

whence

$$(40) \quad m_{ijk} = n_{ijk}(m_{ij.}/n_{ij.}).$$

This adjustment is thus proportionate by tubes, like that in eq. (31), though here the factor  $m_{ij.}/n_{ij.}$  is known and eq. (40) can be applied at once.

*Case V: One set of face totals, and one set of slice totals known.* Sometimes, in addition to the  $rs$  face totals of Case IV, the slice totals

$$N_{..1}, N_{..2}, \dots, N_{..t}$$

will also be known, in which circumstances the conditions (37) are to be accompanied by

$$(41) \quad \sum_{ij} m_{ijk} = m_{..k} = N_{..k} n/N, \quad k = 1, 2, \dots, t-1.$$

The same procedure as previously applied yields now

$$(42) \quad m_{ijk} = n_{ijk}(1 + \lambda_{ij.} + \lambda_{..k})$$

with  $\lambda_{..i}$  to be counted zero. Summations performed over  $k$ , and then over  $i$  and  $j$  together, give the normal equations

$$(43) \quad \begin{aligned} n_{ij} \lambda_{ij} + \sum_k n_{ijk} \lambda_{..k} &= m_{ij} - n_{ij}, \\ \sum_{ij} n_{ijk} \lambda_{ij} + n_{..k} \lambda_{..k} &= m_{..k} - n_{..k}. \end{aligned}$$

The number of equations is  $rs + t - 1$ , since  $\lambda_{..t}$  does not exist. As before, a simplification can be effected by solving the top row for  $\lambda_{ij}$  and making a substitution into the lower one, but because of the great advantage of the iterative process to be seen further on, we shall not carry out the reduction.

Before going on it might be noted that although this case is three dimensional, it reduces to the two dimensional Case II if one considers that  $ij$  is one index running through the values 11, 12,  $\dots$ , 21, 22,  $\dots$ ,  $rs$ , and that  $..k$  is a second index running through the values 1, 2,  $\dots$ ,  $t$ . This can be seen by the similarity between eqs. (43) and (20).

*Case VI: Two sets of face totals known.* If in addition to the face totals of Case IV, the face totals

$$N_{.11}, N_{.12}, \dots, N_{.st}$$

are also known from further crossing the  $j$  and  $k$  characters in the universe, we shall require

$$(44) \quad \sum_i m_{ijk} = m_{.jk} = N_{.jk} n/N, \quad \begin{aligned} j &= 1, 2, \dots, s, \\ k &= 1, 2, \dots, t-1 \end{aligned}$$

in addition to the conditions (37). In place of eq. (40) of Case IV we now find that

$$(45) \quad m_{ijk} = n_{ijk}(1 + \lambda_{ij} + \lambda_{.jk})$$

in which  $\lambda_{.jt}$  is to be counted zero for all  $j$ . No simple relation such as eq. (40) is possible here, because the adjustment is not proportionate by tubes; the Lagrange multipliers must be evaluated. This can be accomplished by summing the members of eq. (45) over  $k$  and  $i$  in turn, resulting in the normal equations

$$(46) \quad \begin{aligned} n_{ij} \lambda_{ij} + \sum_k n_{ijk} \lambda_{.jk} &= m_{ij} - n_{ij}, \\ \sum_i n_{ijk} \lambda_{ij} + n_{.jk} \lambda_{.jk} &= m_{.jk} - n_{.jk}. \end{aligned}$$

Since  $\lambda_{.jt}$  does not exist for any values of  $j$ , the number of equations is  $rs + s(t-1) = s(r+t-1)$ . They break up at once into  $s$  sets each of  $r+t-1$  equations, one set for every  $j$  value. In fact, the problem can be considered as  $s$  sets of the two dimensional Case II. Any one value of  $j$  gives a slice, which can be looked upon as fulfilling the specifications of the two dimensional Case II. Each set of normal equations can be reduced in the same manner that eqs. (20) were reduced.

*Case VII: All three sets of face totals known.* All totals now being known, we require

$$\begin{aligned}
 (37) \quad \sum_k m_{ijk} &= m_{ij.} = N_{ij.} n/N, & i &= 1, 2, \dots, r, \\
 & & j &= 1, 2, \dots, s, \\
 (44) \quad \sum_i m_{ijk} &= m_{.jk} = N_{.jk} n/N, & j &= 1, 2, \dots, s, \\
 & & k &= 1, 2, \dots, t-1, \\
 (47) \quad \sum_j m_{ijk} &= m_{i.k} = N_{i.k} n/N, & i &= 1, 2, \dots, r-1, \\
 & & k &= 1, 2, \dots, t-1.
 \end{aligned}$$

The adjusting relation is

$$(48) \quad m_{ijk} = n_{ijk}(1 + \lambda_{ij.} + \lambda_{.jk} + \lambda_{i.k})$$

in which  $\lambda_{.jt}$  is to be counted zero for any  $j$ ,  $\lambda_{r.k}$  for any  $k$ , and  $\lambda_{i.t}$  for any  $i$ . The normal equations for the Lagrange multipliers are

$$\begin{aligned}
 n_{ij.} \lambda_{ij.} + \sum_k n_{ijk} \lambda_{.jk} + \sum_k n_{ijk} \lambda_{i.k} &= m_{ij.} - n_{ij.} \\
 (49) \quad \sum_i n_{ijk} \lambda_{ij.} + n_{.jk} \lambda_{.jk} + \sum_i n_{ijk} \lambda_{i.k} &= m_{.jk} - n_{.jk} \\
 \sum_j n_{ijk} \lambda_{ij.} + \sum_j n_{ijk} \lambda_{.jk} + n_{i.k} \lambda_{i.k} &= m_{i.k} - n_{i.k}
 \end{aligned}$$

being  $rs + rt + st - r - s - t + 1$  in number. They can be reduced in the same way that previous normal equations have been reduced; but here again, the iterative process will render the use of normal equations unnecessary, except for theoretical purposes, e.g. justification of the iterative process.

**5. A simplified procedure—iterative proportions.** It is well known in least squares that the number of Lagrange multipliers in any problem is equal to the number of conditions imposed on the adjustment. Here the conditions have appeared in sets, depending on which marginal totals are involved. By a comparison of eqs. (15) and (29) on the one hand, with eqs. (19), (31), (34), (42), (45), and (48) on the other, we see that wherever there was only one set of marginal totals involved we came out with a proportionate adjustment, but that in all other cases it was not so; the Lagrange multipliers involved were unfortunately related to one another through normal equations. We now make the observation, however, that as a first approximation the adjustments may all be considered proportionate, and we shall be able to write down an expression for the error in this approximation, and shall be able to eliminate it by a succession of proportionate adjustments.

Take the two dimensional Case II for an example. In eq. (21) one may recognize  $(1/n_{i.}) \sum_j n_{ij.} \lambda_{.j}$  as a weighted average of  $\lambda_{.j}$  for the  $i$ th row. There will be a weighted average of  $\lambda_{.j}$  for the first row, another for the second, etc., one for each value of  $i$ ; consequently one may appropriately speak of the  $i$ th

average of  $\lambda_{.j}$ , writing it  $i\text{-av. } \lambda_{.j}$ . Substituting from eq. (21) into (19) one then sees the adjustment (19) appear as

$$(50) \quad m_{ij} = n_{ij}(m_{i.}/n_{i.} + \lambda_{.j} - i\text{-av. } \lambda_{.j}).$$

If, on the other hand,  $\lambda_{.j}$  had been eliminated from eqs. (20), instead of  $\lambda_{i.}$ , the result would have been

$$(51) \quad m_{ij} = n_{ij}(m_{.j}/n_{.j} + \lambda_{i.} - j\text{-av. } \lambda_{i.}).$$

From either eq. (50) or (51) it is clear why the adjustment (19) is not proportionate by rows or columns, and why Case II does not break up into  $r$  or  $s$  sets of Case I: the reason is that  $\lambda_{.j}$  in any cell is not necessarily equal to the average  $\lambda_{.j}$  for that row, nor is  $\lambda_{i.}$  in any cell necessarily equal to the average  $\lambda_{i.}$  for that column. If nevertheless one were to make the simple proportionate adjustment

$$(52) \quad m'_{ij} = n_{ij}(m_{i.}/n_{i.})$$

along the horizontal in the  $i$ th row, the horizontal conditions (4) will be enforced but not the vertical ones (5); i.e., it will be found that  $m'_{i.} = m_{i.}$ , but that usually not all  $m'_{.j} = m_{.j}$ . This is because eq. (52) effects only a partial adjustment, each  $m'_{ij}$  being in error through the disparity between the  $\lambda_{.j}$  proper to the  $j$ th column, and the average of all the  $\lambda_{.j}$  for the  $i$ th row, as seen in eq. (50). This error can then be diminished by turning the process around and subjecting these  $m'_{ij}$  to a proportionate adjustment in the vertical according to the equation

$$(53) \quad m''_{ij} = m'_{ij}(m_{.j}/m'_{.j})$$

which may be considered an application of eq. (51) wherein the disparity between any  $\lambda_{i.}$  and the average  $\lambda_{i.}$  for the  $j$ th column has been neglected. It is the vertical conditions that will now be found satisfied, but perhaps not all of the horizontal ones, because some of the row totals may have been disturbed. The cycle initiated by eq. (52) is therefore repeated, and the process is continued until the table reproduces itself and becomes rigid with the satisfaction of all the conditions, both horizontal and vertical. The final results coincide with the least squares solution, which is thus accomplished without the use of normal equations.

Usually two cycles suffice. In practice the work proceeds rapidly, requiring only about one-seventh as much time as setting up the normal equations and solving them. The tables III–V show the various stages of the work when the method of iterative proportions is applied to the sample frequencies of Table I. It will be noticed that the results of the third approximation (Table V) are final, since if the process were continued, the table would only reproduce itself.

The same process can be extended to three or more dimensions with an even greater relative saving in time. To see how the method of iterative proportions

applies in one of the three dimensional cases, we may go back to Case III. By the substitution afforded through eq. (36) the adjusting eq. (34) may be put into the form

TABLE III

*The method of iterative proportions applied to the data of Table I. First stage:  
A proportionate adjustment by rows by eq. (52). Note that  $m'_{i.} = m_{i.}$ ,  
but that  $m'_{.j} \neq m_{.j}$*

	$j = 1$	2	3	4	$m'$	$m_{.}$
$i = 1$	3608	778	555	312	5253	5252
2	1586	399	254	157	2396	2395
3	1606	433	273	120	2432	2432
4	10476	2441	1696	1153	15766	15766
5	1660	349	169	152	2330	2330
6	3910	863	548	341	5662	5662
$m'_{.j}$	22846	5263	3495	2235	33839	
$m_{.j}$	22877	5285	3462	2213		33837

TABLE IV

*A continuation of the process initiated in Table III. The figures in Table III are now adjusted proportionately by columns according to eq. (53). The vertical totals  $m''_{.j}$  and  $m_{.j}$  now are equal, but the agreement of the horizontal totals accomplished in Table III has been slightly disturbed*

	$j = 1$	2	3	4	$m''_{.}$	$m_{.}$
$i = 1$	3613	781	550	309	5253	5252
2	1588	401	252	155	2396	2395
3	1608	435	270	119	2432	2432
4	10490	2451	1680	1142	15763	15766
5	1662	350	167	151	2330	2330
6	3915	867	543	338	5663	5662
$m''_{.j}$	22876	5285	3462	2214	33837	
$m_{.j}$	22877	5285	3462	2213		33837

$$(54) \quad m_{ijk} = n_{ijk}(m_{i..}/n_{i..} + \lambda_{.j} + \lambda_{..k} - i\text{-av. } \lambda_{.j} - i\text{-av. } \lambda_{..k}).$$

Equally well it could have been written

$$(55) \quad m_{ijk} = n_{ijk}(m_{.j}/n_{.j} + \lambda_{i..} + \lambda_{..k} - j\text{-av. } \lambda_{i..} - j\text{-av. } \lambda_{..k}),$$

or



$$(56) \quad m_{ijk} = n_{ijk}(m_{..k}/n_{..k} + \lambda_{i..} + \lambda_{.j} - k\text{-av. } \lambda_{i..} - k\text{-av. } \lambda_{.j}).$$

Any of these three equations shows why the adjustment (34) is not proportional by slices, and why this case does not break up into  $r$  or  $s$  or  $t$  sets of the three dimensional Case I. As a first approximation it does, as is now clear from these three equations, and by making successive proportionate adjustments we may thus arrive at the least squares values. To go about the work we could first calculate the values of

$$(57) \quad m'_{ijk} = n_{ijk}(m_{i..}/n_{i..})$$

then

$$(58) \quad m''_{ijk} = m'_{ijk}(m_{.j.}/m'_{.j.})$$

TABLE V

*The cycle is commenced again. The figures of Table IV are subjected to a proportionate adjustment by rows, according to eq. (52). And since these results turn out to be almost a reproduction of Table IV but with both horizontal and vertical conditions satisfied, they are considered final. The agreement with the  $m_{ij}$  in Table I should be noted*

	$j = 1$	2	3	4	$m'_{i.}$	$m_{i.}$
$i = 1$	3612	781	550	309	5252	5252
2	1587	401	252	155	2395	2395
3	1608	435	270	119	2432	2432
4	10492	2451	1680	1142	15765	15766
5	1662	350	167	151	2330	2330
6	3914	867	543	338	5662	5662
$m'_{ij}$	22875	5285	3462	2214	33836	
$m_{ij}$	22877	5285	3462	2213		33837

followed by

$$(59) \quad m'''_{ijk} = m''_{ijk}(m_{..k}/m''_{..k}).$$

These three successive adjustments would constitute a cycle, which would then be repeated in whole or in part until the table becomes rigid with the satisfaction of all three sets of conditions.

**6. Simplification when only one cell requires adjustment.** On occasions it happens in sampling work that one is especially interested in one particular cell of the universe, and would like to have a result for it in advance before the other cells are adjusted. Sometimes it even happens that the others individually are of no particular concern. In such circumstances one merely places the cell

of interest in one corner of the table by an appropriate interchange of rows and columns, and then compresses the rest of the table into the cells adjacent to it. In the two dimensional Case II one would thus work with a  $2 \times 2$  table, one corner cell being the one of special interest, the other three being the result of compression. The marginal totals of the row and column belonging to the cell of interest are unaffected. For illustration we may suppose that from the sample shown in Table I we require only  $m_{61}$ . We then start with the  $2 \times 2$  Table VI, which is derived from Table I by compression. Commencing with Table VI, one might first adjust by rows according to eq. (52), then by columns by eq. (53). One cycle of iterative proportions is sufficient, as is seen in Table

TABLE VI

*Derived from Table I by compression, the cell  $i = 6, j = 1$ , requiring adjustment*

	$j = 1$	$j = 2 - 4$	$n_{i.}$	$m_{i.}$
$i = 1 - 5$	18965	9250	28215	28175
$i = 6$	3882	1740	5622	5662
$n_{.j}$	22847	10990	33837	
$m_{.j}$	22877	10960		33837

TABLE VII

*A proportionate adjustment of Table VI*

Rows adjusted by eq. (52)			Columns adjusted by eq. (53)		
18938	9237	28175	18962	9213	28175
3910	1752	5662	3915	1747	5662
22848	10989	33837	22877	10960	33837

Conclusion:  $m_{61} = 3915$

VII, and the value 3915 found for  $m_{61}$  is in good agreement with its value shown in Tables I and V. The scheme of compression provides a quick method of getting out an advance adjustment for a cell of special interest, and the result so obtained will ordinarily be in good agreement with what comes later when and if all the cells are adjusted.

In the three dimensional Cases II, III, V, VI, and VII, one compresses the original table to a  $2 \times 2 \times 2$  table, and then uses the method of iterative proportions. (The other cases do not require consideration, since they are proportionate adjustments wherein one is already at liberty to adjust as few or as many cells as he likes without altering the equations or the routine.) The same procedure can be extended to the adjustment of two cells, the only modification

being that in two dimensions we shall compress to a  $2 \times 3$  or a  $3 \times 3$  table, depending on whether the two cells do or do not lie in the same row or column. In three dimensions we compress to a  $2 \times 2 \times 3$ , or a  $2 \times 3 \times 3$ , or a  $3 \times 3 \times 3$  table; the first if the two cells lie in the same  $i, j$ , or  $k$  tube, the second if they lie in the same slice but not in the same tube, the third if they are in separate slices.

**7. Some remarks on the accuracy of an adjustment.** A least squares adjustment of sampling results must be regarded as a systematic procedure for obtaining satisfaction of the conditions imposed, and at the same time effecting an improvement of the data in the sense of obtaining results of smaller variance than the sample itself, under ideal conditions of sampling from a stable universe. It must not be supposed that any or all of the adjusted  $m_{ij}$  in any table are necessarily "closer to the truth" than the corresponding sampling frequencies  $n_{ij}$ , even under ideal conditions. As for the standard errors of the adjusted results, they can easily be estimated for the ideal case by making use of the calculated chi-square. For predictive purposes, however (which can be regarded as the only possible use of a census by any method, sample or complete), it is far preferable, in fact necessary, to get some idea of the errors of sampling by actual trial, such as by a comparison of the sampling results with the universe, as can often be arranged by means of controls. There is another aspect to the problem of error—even a 100 per cent count, even though strictly accurate, is not by itself useful for prediction, except so far as we can assert on other grounds what secular changes are taking place.

In conclusion it is a pleasure to record our appreciation of the assistance of Miss Irma D. Friedman and Mr. Wilson H. Grabill for putting the formulas and procedure into actual operation with census data, and thereby disclosing defects in earlier drafts of the manuscript.

BUREAU OF THE CENSUS,  
WASHINGTON

## NOTES

*This section is devoted to brief research and expository articles, notes on methodology and other short items.*

### THE STANDARD ERRORS OF THE GEOMETRIC AND HARMONIC MEANS AND THEIR APPLICATION TO INDEX NUMBERS<sup>1</sup>

BY NILAN NORRIS

Attempts to derive useful expressions for estimating the standard deviations of the sampling errors of the geometric and harmonic means have not yielded results comparable with those afforded by the modern theory of estimation, including fiducial inference. There are in the literature of probability theory certain theorems which can be applied to obtain these desired results in a straightforward manner. The use of forms for estimating standard errors is subject to certain conditions which are not always fulfilled, particularly in the case of time series. An understanding of these limitations should deter those who may be tempted to judge the significance of phenomena such as price changes solely on the basis of estimated standard errors of indexes.

**1. Statement of formulas.** The standard error of the geometric mean of a sequence of positive independent chance variables denoted by  $x_i = x_1, x_2, \dots, x_n$ , is  $\sigma_G = \theta_1 \frac{\sigma_{\log x}}{\sqrt{n}}$ , where  $\theta_1$  is the population geometric mean of the variates; so that  $\sigma_{\log x}$  is the standard deviation of the logarithms in the population as given by  $\sigma_{\log x} = [E\{[\log x - E(\log x)]^2\}]^{1/2}$ ; and  $n$  is the number of individuals comprising the sample. The estimate of the standard error of the geometric mean is  $s_G = G \frac{s_{\log x_i}}{\sqrt{n-1}}$ , where  $G$  is the sample geometric mean, that is, the estimate of  $\theta_1$ ; so that  $s_{\log x_i}$  is the estimate of  $\sigma_{\log x}$ ; and  $n-1$  is the degree of freedom of the sample.

<sup>1</sup> This article summarizes two papers presented at sessions of the Institute of Mathematical Statistics at Detroit, Michigan on December 27, 1938, and at Philadelphia, Pennsylvania on December 27, 1939. The results given herein can be derived by several methods, which vary somewhat as to degree of rigor. The writer wishes to acknowledge his indebtedness to the referee for suggesting a proof based on a probability theorem stated by J. L. Doob, "The limiting distributions of certain statistics," *Annals of Math. Stat.*, Vol. 4 (1935), pp. 160-169. The standard deviation formulas obtained follow as an application of this theorem, as will be seen by reference to it. Obviously the asymptotic variance formulas of many other statistics (estimates of parameters) can be obtained in a similar manner.

The standard error of the harmonic mean of a sequence of positive independent chance variables denoted by  $x_i = x_1, x_2, \dots, x_n$ , is  $\sigma_H = \theta_2^2 \frac{\sigma_{1/x}}{\sqrt{n}}$ , where the population harmonic mean of the variates is  $\theta_2 = 1/\alpha = [E(1/x)]^{-1}$ ; so that the standard deviation of  $1/x$  in the population is  $\sigma_{1/x} = [E\{[1/x - E(1/x)]^2\}]^{1/2}$ ; and  $n$  is the number of observations comprising the sample. The estimate of the standard error of the harmonic mean is  $s_H = \frac{1}{a^2} \frac{s_{1/x_i}}{\sqrt{n-1}}$ , where the estimate of  $\alpha$  is given by  $a = \frac{1}{H} = \frac{1}{n} (\sum 1/x_i)$ ; in which  $s_{1/x_i}$  is the standard deviation of the reciprocals of the observations comprising the sample; and  $n - 1$  is the degree of freedom of the sample.

**2. Derivation of formulas.** These forms can be obtained by application of the Laplace-Liapounoff theorem<sup>2</sup> as follows: Let  $x_i = x_1, x_2, \dots, x_n$  be a set of positive independent chance variables with the same distribution functions, where the expectations,  $E(x_i)$  and  $E(x_i^2)$  exist, and where  $\sigma_x^2 = E\{[x_i - E(x_i)]^2\} > 0$ . The last condition is imposed to eliminate the trivial case in which the  $x_i$  are all equal and their distribution is confined to a single point. The geometric mean of the  $x_i$  is  $G = (x_1 \cdot x_2 \cdot \dots \cdot x_n)^{1/n}$ , and the harmonic mean of the  $x_i$  is  $H = \left[ \frac{1}{n} \sum \frac{1}{x_i} \right]^{-1}$ .

It is necessary to assume that both  $\sigma_{\log x}^2$  and  $\sigma_{1/x}^2$  are finite, and that in the case of both  $\log x$  and  $1/x$  at least one moment of order higher than any two of the respective variates is also finite. The requirement that the variance and at least one moment higher than the variance be finite can be weakened in various ways, but this is a trivial consideration, since nearly all distributions of any importance have finite third moments.<sup>3</sup> Certain rarely occurring types of distributions, such as the Cauchy distribution, have infinite variance. In such cases, standard error formulas as ordinarily used are not valid.

Let  $E(\log x) = \zeta$ , and  $E(1/x) = \alpha$ . By the Laplace-Liapounoff theorem, except for terms of order  $1/\sqrt{n}$ , the limiting distributions of  $\frac{\sqrt{n}(\log G - \zeta)}{\sigma_{\log x}}$  and  $\frac{\sqrt{n}(H^{-1} - \alpha)}{\sigma_{1/x}}$  are normal with zero arithmetic means and unit variances. That is, if  $C$  represents a set of conditions on chance variables, and  $P\{C\}$  is the probability that these conditions are satisfied, then

<sup>2</sup> A. Khintchine, *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, J. Springer, Berlin, 1933, Vol. II, No. 4, pp. 1-8; J. L. Doob, *op. cit.*, pp. 160-169; and S. S. Wilks, *Statistical Inference*, 1936-1937, Edwards Brothers, Inc., Ann Arbor, 1937, pp. 39 f.

<sup>3</sup> For a more detailed discussion of this matter see Wilks, *op. cit.*, pp. 39 f.

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\sqrt{n}(\log G - \zeta)}{\sigma_{\log G}} < t \right\} = \lim_{n \rightarrow \infty} P \left\{ \frac{\sqrt{n}(H^{-1} - \alpha)}{\sigma_{1/n}} < t \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx.$$

In order to use these relations in obtaining the limiting distributions of the geometric and harmonic means, it is necessary to suppose that the sequence of random chance variables,  $V_i$ , converges in probability (converges stochastically) to  $\rho$ , and that the sequence of random chance variables,  $\sqrt{n}(V_i - \rho)$ , has a normal limiting distribution with zero arithmetic mean and variance  $\sigma^2$ . Also, it is necessary to assume that the real-valued function,  $f(x)$ , has a Taylor expansion valid in the neighborhood of  $\rho$ . If  $f'(\rho) \neq 0$ , only the first two terms of the series are needed. The required expansion is given by

$$f(x) = f(\rho) + (x - \rho)f'(\rho) + \frac{(x - \rho)^2}{2} f''[\rho + \beta(x - \rho)],$$

where  $0 < \beta < 1$ , and  $f''(x)$  is continuous in the neighborhood of  $\rho$ . When these conditions are fulfilled, the limiting distribution of  $\sqrt{n}[f(V_i) - f(\rho)]$  is normal with an arithmetic mean of zero and a variance of  $\sigma^2[f'(\rho)]^2$ .

Let  $f(\log G) = e^{\log G}$ , and use the expansion given by  $e^{\log G} = e^{\zeta} + (\log G - \zeta)e^{\zeta} + \frac{1}{2}(\log G - \zeta)^2 e^{\zeta + \beta(\log G - \zeta)}$ . Since  $\theta_1 = e^{\zeta}$ , it follows that the limiting distribution of  $\sqrt{n}(G - \theta_1)$  is normal with an arithmetic mean of zero and a variance of  $\theta_1^2 \sigma_{\log G}^2$ .

Similarly, it can be shown that the limiting distribution of  $\sqrt{n}(H - \theta_2)$  is normal with an arithmetic mean of zero and a variance of  $\theta_2^4 \sigma_{1/x}^2$ , where  $\theta_2 = \frac{1}{\alpha} = [E(1/x)]^{-1}$ .

It is of some interest to observe that the expressions for the standard errors of the geometric and harmonic means correspond with the forms previously given for the standard errors of two efficient ratio-measures of relative variation,<sup>4</sup> namely,

$$\sigma_{G/A} = \frac{\theta_1^2}{\theta^2} \sigma_{A/G}, \quad \text{and} \quad \sigma_{H/G} = \frac{\theta_2^2}{\theta_1^2} \sigma_{G/H},$$

where  $\theta_1/\theta$  is the population geometric-arithmetic ratio, and  $\theta_2/\theta_1$  is the population harmonic-geometric ratio.

**3. Limitations of standard-error estimates.** Application of these forms is subject to the usual conditions for drawing sound inferences on the basis of the representative method. Fiducial argument should be employed to avoid certain untenable assumptions of the outmoded method of using standard errors. Estimates of the standard deviations of sampling errors do not constitute an ultimate test of significance which can be applied with a high degree of success to all types of problems. In general, such estimates cannot be relied upon with a

<sup>4</sup> Nilan Norris, "Some efficient measures of relative dispersion," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 214-220.

high degree of confidence when they are used as tests of significance for index numbers, since in nearly all time series there exists an appreciable degree of serial correlation, persistence, or lack of independence among successive items of any sample.

**4. Bibliographical note.** Certain aspects of the sampling distribution of the geometric mean have been discussed by Burton H. Camp.<sup>6</sup> Attempts to derive forms for estimating the standard errors of index numbers have been made by Truman L. Kelley<sup>6</sup> and Irving Fisher,<sup>7</sup> and an empirical study of the sampling fluctuations of indexes has been made by E. C. Rhodes.<sup>8</sup> Although various special tests of significance for time series have been proposed,<sup>9</sup> at the present time no generally satisfactory procedure has appeared.

HUNTER COLLEGE,  
NEW YORK, N. Y.

<sup>6</sup> Burton H. Camp, "Notes on the distribution of the geometric mean," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 221-226.

<sup>6</sup> Truman L. Kelley, "Certain Properties of Index Numbers," *Quarterly Publications of Am. Stat. Assn.*, Vol. 17, New Series 135, Sept., 1921, pp. 826-841.

<sup>7</sup> Irving Fisher, *The Making of Index Numbers*, Houghton Mifflin Company, New York, 1927, 3d ed., pp. 225-229, 342-345, and Appendix I, pp. 407 and 430 f.

<sup>8</sup> E. C. Rhodes, "The precision of index numbers," *Roy. Stat. Soc. Jour.*, Vol. 99 (1936), Part I, pp. 142-146, and Part II, pp. 367-369.

<sup>9</sup> Some of the more recent papers dealing with this matter are: G. Tintner, "On tests of significance in time series," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 139-143; "The analysis of economic time series," *Am. Stat. Assn. Jour.*, Vol. 35 (1940), pp. 93-100; L. R. Hafstad, "On the Bartels technique for time-series analysis, and its relation to the analysis of variance," *Am. Stat. Assn. Jour.*, Vol. 35 (1940), pp. 347-361; and Lila F. Knudsen, "Interdependence in a series," *Am. Stat. Assn. Jour.*, Vol. 35 (1940), pp. 507-514.

## A NOTE ON THE USE OF A PEARSON TYPE III FUNCTION IN RENEWAL THEORY

By A. W. BROWN

One of the methods suggested by A. J. Lotka<sup>1</sup> for the derivation of the renewal function may be briefly summarized as follows.

The method consists of dissecting the total renewal function into "generations". The original installation constitutes the zero generation, the units introduced to replace disused units of the zero generation constitute the first generation, renewal of these the second, and so on. Let  $f(x)$  be the "mortality" function, the same for all generations.  $f(x)$  is a function satisfying the usual conditions of a distribution function. Adopting Lotka's notation, let  $N$  be the number of units in the original collection,  $B_1(t) dt$  the number of objects intro-

<sup>1</sup> A. J. Lotka, "A Contribution to the Theory of Self Renewing Aggregates, With Special Reference to Industrial Replacement," *Annals of Math. Stat.*, Vol. 10 (1939), p. 1.

duced between times  $t$  and  $t + dt$  and belonging to the first generation,  $B_1(t) dt$  a similar expression for the second generation, etc.  $B_1(t)/N$ ,  $B_2(t)/N$ , ... may be regarded as renewal density functions for the various generations.

Now, evidently,

$$(1) \quad B_1(t) = Nf(t)$$

$$(2) \quad B_2(t) = \int_0^t B_1(t-x)f(x) dx$$

and in general

$$(3) \quad B_{i+1}(t) = \int_0^t B_i(t-x)f(x) dx.$$

Summation of the contributions of the successive generations gives for the total renewal at the time  $t$

$$(4) \quad B(t) = B_1(t) + \int_0^t B(t-x)f(x) dx.$$

In this note we propose to use a Pearson Type III function for  $f(x)$  and observe what form our equations then assume. The Pearson Type III function  $\frac{c^k}{\Gamma(k)} x^{k-1} e^{-cx}$ , ( $c > 0$ ,  $k > 0$ ), appears to be a reasonable one to use in many practical situations. The two parameters  $c$  and  $k$  give it a considerable amount of flexibility. The fact that this function has an unlimited range in one direction is relatively unimportant from a practical point of view, as is well known from the experience of fitting curves of this type to skewed data with limited range. Of course the question of whether a Type III curve is appropriate can be answered more objectively by using the usual Pearson curve-fitting criteria,  $\beta_1$ ,  $\beta_2$  and  $k$ . We have, then, substituting in (1)

$$(5) \quad B_1(t) = N \frac{c^k}{\Gamma(k)} t^{k-1} e^{-ct}$$

and from (2)

$$(6) \quad B_2(t) = \int_0^t N \frac{c^k}{\Gamma(k)} (t-x)^{k-1} e^{-c(t-x)} \frac{c^k}{\Gamma(k)} x^{k-1} e^{-cx} dx$$

$$(7) \quad = \frac{Nc^{2k}}{\Gamma(k)\Gamma(k)} e^{-ct} \int_0^t (t-x)^{k-1} x^{k-1} dx.$$

If, now, we set  $x = ty$ , the integral in (7) reduces to

$$\int_0^1 (t-y)^{k-1} y^{k-1} dy = t^{2k-1} \frac{\Gamma(k)\Gamma(k)}{\Gamma(2k)}.$$



Hence,

$$(8) \quad B_2(t) = N \frac{c^{2k}}{\Gamma(2k)} t^{2k-1} e^{-ct}$$

and in general

$$(9) \quad B_j(t) = N \frac{c^{jk}}{\Gamma(jk)} t^{jk-1} e^{-ct}.$$

Summing the contributions of the several generations, we have for the total renewal function

$$(10) \quad B(t) = Nce^{-ct} \left\{ \frac{(ct)^{k-1}}{\Gamma(k)} + \frac{(ct)^{2k-1}}{\Gamma(2k)} + \dots \right\}.$$

If  $k$  is a positive integer  $\geq 3$ , (10) can be easily summed to a form which shows immediately its damped periodic nature. Even if  $k$  is positive but not an integer, it can be shown by continuity considerations that the function  $B(t)$  defined by (10) has periodic properties.

Assuming  $k$  to be a positive integer, then, and setting  $z = ct$ , we may write the expression in brackets in (10) as

$$(11) \quad \frac{z^{k-1}}{(k-1)!} + \frac{z^{2k-1}}{(2k-1)!} + \dots = f(z).$$

Then

$$\frac{d^k f(z)}{dz^k} = f(z)$$

and upon making the trial substitution,  $f(z) = Ae^{ms}$ , we get

$$Am^k e^{ms} = Ae^{ms}.$$

Hence,

$$m^k = 1.$$

Taking unity in its complex form

$$1 = \cos 2n\pi + i \sin 2n\pi$$

we have that

$$(12) \quad m_n = \sqrt[k]{1} = \cos \frac{2n\pi}{k} + i \sin \frac{2n\pi}{k}$$

where  $n = 0, 1, 2, \dots, k-1$ . Then

$$f(z) = \sum_{n=0}^{k-1} A_n e^{m_n z}$$

and

$$f'(z) = \sum_{n=0}^{k-1} A_n m_n^i e^{m_n z}.$$

Now setting  $z = 0$ , we get

$$f(0) = A_0 + A_1 + \dots + A_{k-1} = 0$$

$$f'(0) = A_0 m_0 + A_1 m_1 + \dots + A_{k-1} m_{k-1} = 0$$

$$f^{k-1}(0) = A_0 m_0^{k-1} + A_1 m_1^{k-1} + \dots + A_{k-1} m_{k-1}^{k-1} = 1$$

$k$  equations to determine the  $k$  constants. We know that  $A_n$  is equal to the ratio of two determinants formed from the coefficients of the above equations. This ratio reduces to

$$(13) \quad A_n = \frac{(-1)^{k+n+1}}{(m_{k-1} - m_n)(m_{k-2} - m_n) \dots (m_n - m_0)}.$$

We have, then, an expression for the  $k$  constants in terms of the  $k$  roots of unity. Therefore, for any particular value of  $k$  we can obtain the sum of our series from the relation

$$f(z) = \sum_{n=0}^{k-1} A_n e^{m_n z}.$$

Hence, under the assumption that  $k$  is a positive integer, we have

$$(14) \quad B(t) = Nce^{-ct} \sum_{n=0}^{k-1} A_n e^{m_n ct}.$$

The forms of  $B(t)$  for  $k = 1, 2, 3, 4$  are respectively

$$B(t) = Nc$$

$$B(t) = \frac{1}{2}Nc(1 - e^{-2ct})$$

$$B(t) = Nce^{-ct} \left[ \frac{1}{2}e^{ct} - e^{-ct} \left( \frac{1}{2} \cos \frac{1}{2}\sqrt{3}ct + \frac{1}{\sqrt{3}} \sin \frac{1}{2}\sqrt{3}ct \right) \right]$$

$$B(t) = Nce^{-ct} \left[ \frac{1}{4}(e^{ct} - e^{-ct}) - \frac{1}{2} \sin ct \right].$$

Although the above procedure is valuable particularly because it brings to light something of the nature of our renewal function, the forms derived above can be used actually to obtain values of  $B(t)$  for various values of  $t$ . However, for extensive numerical work a better method is at hand, which does not even depend on the assumption of an integral value for  $k$ .

Let us return once again to equation (10) which may be written in the following form

$$(15) \quad B(t) = Nc \left\{ \frac{e^{-ct}(ct)^{k-1}}{\Gamma(k)} + \frac{e^{-ct}(ct)^{2k-1}}{\Gamma(2k)} + \dots \right\}.$$

If  $k$  and  $c$  are determined by the method of moments, (using two moments),  $k$  will not, in general, be a positive integer. However, by using the *Tables of the Incomplete Gamma Function* edited by Karl Pearson, one can compute values of  $B(t)$  without much difficulty. In these tables the function  $I(u, p)$  is tabulated for various values of  $u$  and  $p$ , where  $I(u, p)$  is defined by

$$(16) \quad I(u, p) = \frac{\int_0^{u\sqrt{p+1}} e^{-v} v^p dv}{\Gamma(p+1)}.$$

If we let  $\xi = u_1\sqrt{p+1} = u_0\sqrt{p}$  then upon integrating by parts we find

$$(17) \quad \frac{e^{-\xi}\xi^p}{\Gamma(p+1)} = I(u_0, p-1) - I(u_1, p).$$

The left hand member of this equation is of the same form as each of the terms of the series in brackets in (15). Hence, the value of the renewal function for a particular time,  $t$ , is directly obtainable by summation of the right hand member of (17) for successive significant values of the argument  $p$ .

By way of illustration a numerical example will be considered. The data are taken from E. B. Kurtz' book entitled *Life Expectancy of Physical Property*. In this book the author makes a study of retirement rates of fifty-two different types of physical property, and finds that their replacement curves fall into seven distinct groups. We consider here Group VII which happens to be the largest group, embracing seventeen different types of industrial equipment out of the fifty-two examined. Using Kurtz' replacement data <sup>2</sup> we obtain for the value of the first and second moments

$$\mu_1 = 10.002$$

$$\mu_2 = 121.71$$

and from these by the method of moments, we find

$$k = 4.62$$

$$c = .462.$$

We then proceed to calculate values of  $B(t)/N$  by means of Pearson's Tables,<sup>3</sup> obtaining the results shown in the following table.

<sup>2</sup> E. B. Kurtz, *Life Expectancy of Physical Property*, Ronald Press, 1930, Table 22, page 86.

<sup>3</sup> With regard to the method of interpolation employed in the calculations, it should be mentioned that it was found advisable to use the Mid-panel Central Difference Formula (xxiii) on page xii of the introduction to Pearson's Tables; and that it is quite sufficient for our purposes to calculate only first order terms.

$t$	$B(t)/N$	$t$	$B(t)/N$
0	.0000	10	.1049
1	.0016	11	.1043
2	.0103	12	.1028
3	.0279	13	.1006
4	.0486	14	.0990
5	.0714	15	.0994
6	.0867	16	.1009
7	.0980	17	.1013
8	.1039	18	.0992
9	.1066	19	.0999
		20	.0993

In conclusion the author wishes to thank Professor S. S. Wilks for various suggestions he has made in connection with this note.

PRINCETON UNIVERSITY,  
PRINCETON, N. J.

## ESTIMATES OF PARAMETERS BY MEANS OF LEAST SQUARES

BY EVAN JOHNSON, JR.

As a criterion for comparing estimates of a parameter of a universe, of known type of distribution, the use of the principle of least squares is suggested. A criterion may be stated in rather general terms. Its application to any given problem presumes a knowledge of the distribution functions of the estimates considered. In the present paper a criterion is set up and application of it is made in the estimation of the mean and of the square of standard deviation of a normal universe.

We shall use the symbol  $\theta$  to represent a parameter to be estimated. It is to be remembered that  $\theta$  is a constant throughout any problem, that it represents an unknown value, and that observations and functions of observations (called estimates) are the only variables that occur. We shall use the symbols  $x_i$ ,  $i = 1, 2, \dots, n$ , to represent observed values of the variable  $x$  of the universe, and the symbol  $F$  to represent a given function of the observations  $x_i$ .

If we choose to consider a given function  $F$  as an estimate of  $\theta$ , we are then interested in the error  $F - \theta$ . This quantity differs from the so-called residual of least square theory, since we are here interested in the difference between computed and true values, rather than in the difference between observed and computed values. To avoid any possible confusion we shall refer to  $F - \theta$  as the error. Over the set of all samples of  $n$  observations,  $x_i$ , the distribution of the errors  $F - \theta$  is expressed by means of the distribution function  $f(F)$ ,

which may be computed from the known distribution function of the universe. We shall assume that the function  $f(F)$  has been normalized, so that  $\int_{\alpha}^{\beta} f(F) dF = 1$ , where the interval from  $\alpha$  to  $\beta$  includes all possible values of  $F$ . The integral  $I = \int_{\alpha}^{\beta} (F - \theta)^2 f(F) dF$ , associated with a given estimate  $F$ , may be thought of as the average square error over the set of all samples.

In this notation we shall state a criterion for the judgment of estimates in either of the two following forms:

DEFINITION 1. Let  $f_1$  be the distribution function of  $F_1$ , and  $f_2$  that of  $F_2$ . The estimate  $F_1$  of  $\theta$  will be judged better than the estimate  $F_2$  if

$$\int_{\alpha}^{\beta} (x - \theta)^2 f_1(x) dx < \int_{\alpha}^{\beta} (x - \theta)^2 f_2(x) dx.$$

DEFINITION 2. From a given class of functions, of which  $F$  is a member,  $F$  will be called the best estimate if

$$(1) \quad I = \int_{\alpha}^{\beta} (F - \theta)^2 f(F) dF$$

is less than the corresponding integral for all other functions of the class.

It is to be observed that the integral  $I$  is a function of the quantities  $\theta$  and  $f$ . From this is seen at once the distinction between the present problem of minimizing the average square error and the similar problem of finding that point around which the mean square value of the deviations of a variable is a minimum. In the problem under consideration we wish to find the function  $F$ , or more precisely its distribution function  $f(F)$ , for which  $I$  takes its minimum with a fixed value of  $\theta$ . In the alternative problem we have a given distribution  $f$  and we wish to find the minimum of  $I$  with respect to  $\theta$ .

A second observation to be made is that the integral  $I$  can not be usefully minimized in the sense of the general conditions of the calculus of variations. The problem would be of the isoperimetric variety, with the side condition  $\int_{\alpha}^{\beta} f(x) dx = 1$ . A solution might be expressed as the limit, as  $a$  approaches zero, of functions  $f(x)$  with proper continuity conditions, such that

$$f(x) \begin{cases} = 0 & \text{when } |x - \theta| \geq a, \\ > 0 & \text{when } |x - \theta| < a, \text{ and } \int_{\theta-a}^{\theta+a} f(x) dx = 1. \end{cases}$$

Such a solution would be meaningless in practical statistical theory. Solutions are to be expected, therefore, only in those cases where the class of functions, from which  $F$  is to be selected, is sufficiently restricted.

The two following examples illustrate both restrictions and possible application of the theory.

As a first example let us consider the problem of finding an estimate  $F$  of the mean,  $\bar{x}$ , of a normal universe. The mean of a distribution is a symmetric linear function of the variates of the distribution. For the class of functions from which to select an estimate  $F$  of  $\bar{x}$ , let us take the class of all symmetric homogeneous linear functions of the observations  $x_i$ . Let

$$(2) \quad F = a(x_1 + x_2 + \dots + x_n).$$

We wish to find the value of  $a$ , if any, for which  $I$  is a minimum.

$F$  is the sum of  $n$  normally distributed independent variables,  $ax_i$ , each with standard deviation  $a\sigma$ .  $F$ , therefore, has a distribution function

$$f = C \cdot \exp \left( \frac{-(F - an\bar{x})^2}{2a^2n\sigma^2} \right),$$

where  $C$  is so chosen that  $\int_{-\infty}^{\infty} f dF = 1$ . A discussion of general distribution functions may be found in Dunham Jackson's article, "Theory of Small Samples," in the *American Mathematical Monthly*, Volume XLII, 1935. In this case it can be shown without particular difficulty that

$$\begin{aligned} I &= C \int_{-\infty}^{\infty} (F - \bar{x})^2 \cdot \exp \left( \frac{-(F - an\bar{x})^2}{2a^2n\sigma^2} \right) dF \\ &= a^2n\sigma^2 + \bar{x}^2(an - 1)^2. \end{aligned}$$

To determine the minimum of  $I$  with respect to  $a$ , we set

$$\frac{\partial I}{\partial a} = 2an\sigma^2 + 2\bar{x}^2(an - 1)n = 0,$$

and obtain

$$(3) \quad a = \frac{\bar{x}^2}{n\bar{x}^2 + \sigma^2} = \frac{1}{n} \frac{1}{1 + \sigma^2/n\bar{x}^2} \\ = \frac{1}{n} \left( 1 - \frac{\sigma^2}{n\bar{x}^2} + \dots \right).$$

It is seen that for even such a simple example as the estimation of the mean there is no estimate of the form of equation (2), with  $a$  independent of the parameter to be estimated, for which  $I$  takes its minimum value.

For a distribution in which  $\bar{x} \neq 0$ , and  $\sigma^2/n\bar{x}^2$  is small,  $a$  is given as a first approximation by  $1/n$ . The function  $F$  is merely the mean of the sample observations. If  $\bar{x} = 0$ , the required solution is  $a = 0$ , and there is no best least square estimate of the type of equation (2).

In the case where  $\sigma^2/\bar{x}^2$  is not small, as is apt to be the case when  $\bar{x}$  is near zero, the determination of a desirable estimate by least squares requires a knowledge of the ratio  $\sigma^2/\bar{x}^2$ , which may perhaps be judged approximately in a special

problem. If this value is assumed known, the required value of  $a$  may be found most easily by rewriting equation (3) in the form

$$(4) \quad a = \frac{1}{n + \sigma^2/\bar{x}^2}.$$

The second example to be considered is the determination of an estimate of  $\sigma^2$  of a normal universe. A comparison with the definition of  $\sigma^2$  suggests the use of a function  $F$  given by the equation

$$(5) \quad F = a \{ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \},$$

where  $\bar{x}$  is the mean of the  $n$  observations. The value of  $a$  is, of course, to be determined by minimizing the integral  $I$ .

$F$  is the sum of the squares of  $n$  normally distributed but not independent variables. It may be shown, however, (Jackson, *loc. cit.*) to be expressible as the sum of the squares of  $n-1$  independent normally distributed variables, each with standard deviation  $\sqrt{a}\sigma$ . The distribution function for  $F$  takes the form

$$(6) \quad f(F) = C (F)^{(n-3)/2} e^{-F/2a\sigma^2},$$

$F$  taking only positive values, and  $C$  is again chosen to normalize  $f(F)$ . The integral  $I$  may be written

$$I = C \int_0^\infty (F - \sigma^2)^2 (F)^{(n-3)/2} e^{-F/2a\sigma^2} dF.$$

The integration is most easily accomplished by replacing  $F$  by  $u^2$ , and in terms of  $u$

$$I = C' \int_0^\infty (u^2 - \sigma^2)^2 u^{n-2} e^{-u^2/2a\sigma^2} du.$$

The various steps in the integration will differ for even and odd values of  $n$ , but in each case the final result is the same. It is found that

$$(7) \quad I = \sigma^4 \{ a^2(n^2 - 1) - 2a(n - 1) + 1 \}.$$

The value of  $a$  which minimizes  $I$  is determined from the relation

$$\frac{\partial I}{\partial a} = \sigma^4 \{ 2a(n^2 - 1) - 2(n - 1) \} = 0.$$

Dividing by  $(n-1)$ , which is not zero in a sample of two or more observations, we obtain

$$(8) \quad a = \frac{1}{n + 1}.$$

In contrast to the previous example we have here an absolute minimum of  $I$  with respect to all estimates of the type of equation (5). The best least square estimate of this type is, therefore,

$$(9) \quad F = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n + 1}.$$

## THE TEACHING OF STATISTICS<sup>1</sup>

BY HAROLD HOTELLING

The very great increase in the teaching of statistics since the First World War has been associated on one hand with the development of statistical theory. This important series of discoveries has made available more and more powerful and accurate statistical methods, and has also acquired an intellectual interest of its own as embodying the modern version of the most important part of inductive logic and as providing scope for mathematical and logical ingenuity of high order. The increased teaching of statistics has also been associated with the rapidly growing applications of statistics in innumerable fields, made possible by the development of the theory, by the availability of persons having some knowledge of the theory, and by an increasing realization of the possibilities of application. Doubtless most students of statistics enter upon the subject, not for its intrinsic interest, but with the idea of applying statistical methods as a tool to some particular end. This object may be scientific research, or to fulfill a requirement for a degree, but is often connected with some purely practical pursuit offering the ready prospect of a remunerative job. But it would be a mistake to ignore those whose interest is more purely intellectual, who desire an insight into the peculiar problems of probable inference and the structure of empirical knowledge, who wish to get a fundamental acquaintance with one of the most fundamental of subjects, to see and understand fully the mathematical derivations underlying so much practical and scientific activity, and perhaps to make their own contributions.

Of the magnitude of the demand for statisticians there can be no doubt. The realization of what statistical methods can do in a multitude of fields has gradually led the administrators of government agencies, directors of scientific organizations and research institutes, and business men, to employ rapidly increasing numbers of persons with some knowledge of statistical methods, and to accord an unusual degree of recognition and promotion in many such cases. The uses of statistical methods, and especially of sampling theory, are so varied that it is scarcely possible in a brief space to give any sort of survey of them. They enter, in one form or another, into the research work of the physicist, the chemist, the astronomer, the biologist, the psychologist, the anthropologist, the medical investigator, the economist, and the sociologist. Meteorology, which has lately acquired greatly increased importance, both civil and military, is with its masses of numerical observations very much a statistical matter. The engineer needs modern statistical methods both in the physical and in the

<sup>1</sup> Address at the meeting of the Institute of Mathematical Statistics at Hanover, N. H., September 10, 1940.



economic aspects of his plans. The work of W. A. Shewhart has made clear the central importance of sampling theory in the economic control of quality of manufactured articles. Business men who use sampling surveys to test the markets for their products and the effectiveness of their advertising, who employ statisticians to make up index numbers and forecasts of business conditions, and whose manufacturing costs and quality are controlled with the help of recently devised statistical methods, are finding more and more uses for statisticians. Indeed, it seems as if the exploitation of the business and manufacturing possibilities of statistical methods has only begun, and that limitless further fields are coming into view. Insurance has of course always been essentially dependent on statistics.

But the most rapidly growing large class of positions for statisticians is at present in governmental activities. For some facts regarding the employment of statisticians by the federal government I am indebted to Dr. J. M. Thompson. It appears that it has about one hundred agencies using statistics, with almost eight hundred positions broadly classified as statistical or mathematical, in addition to more than six thousand generally classified as economists. The title "economist" covers many types of work, but much of it is largely statistical. The nature of the government's statistical work is varied and extensive. It includes such work as forecasting revenue from taxes, prices and production of agricultural commodities, general demand conditions, and weather. Some of the work consists in analyzing the effects of various taxes on other programs. In connection with proposed legislation, statisticians serving the lawmakers often attempt to outline the probable results of the legislation, as well as to assist in setting up definite formulae for carrying out the general policies aimed at in Acts of Congress. Administrators as well as lawmakers require statistical activities of a high order, exemplified in the Bureau of the Census, the Bureau of Agricultural Economics, and others. The scientific activities of the government, the work of the War Department, and many others that do not at first sight appear at all statistical, require the services of mathematical statisticians of high order. Even the judicial activities call for statistical theory of some of the most recently discovered kinds, as for instance in the investigation recently made of parole procedures. Cities and states, school and port authorities, employ numerous statisticians for other and widely diverse purposes.

The growing need, demand and opportunity have confronted the educational system of the country with a series of problems regarding the teaching of statistics. Should statistics be taught in the department of agriculture, anthropology, astronomy, biology, business, economics, education, engineering, medicine, physics, political science, psychology, or sociology, or in all these departments? Should its teaching be entrusted to the department of mathematics, or to a separate department of statistics, and in either of these cases should other departments be prohibited from offering duplicating courses in statistics, as they are often inclined to do? To what students, and at what stage of their advancement, should a course in statistics be administered?

Should there be mathematical or other prerequisites? How much of an investment in a statistical laboratory is warranted? Should courses be primarily theoretical and mathematical, or should they be made as practical as possible, equipping the student in the shortest possible time for a job as statistician, or for statistical work in the field with which a particular department is concerned? What about degrees in statistics? Eclipsing all these in importance, though it seems to have received too little of the attention of college and university administrative officers is the question, What sort of persons should be appointed to teach statistics?

To pressing practical problems answers are sure to be given either by considered policy or by processes of historical evolution. The latter are the more prominent in explaining the statistical teaching we have had. A synoptic picture of the origins, not many decades ago, of a good deal of it would perhaps be something like this. A university Department of X, where X stands for economics, psychology, or any one of numerous other fields, begins to note toward the end of the pre-statistical era that some of the outstanding work in its field involves statistics. The quantity and importance of such work are observed to increase, while at the same time its intelligibility seems to diminish. Evidently students turned out with degrees in the field of X who do not know something about statistics are going to be handicapped, and are not likely to reflect credit on Alma Mater. The department therefore resolves that its students must acquire at least an elementary knowledge of the fundamentals of statistics. To implement this principle, it perhaps inserts some acquaintance with statistics among the requirements for a degree. This situation naturally calls for the introduction of a course in statistics. Accordingly the head of the Department of X, in preparing the next Announcement of Courses, writes:

"X 82. Elements of Statistics. An elementary but thorough course designed to acquaint students of X with the fundamental concepts of statistics and their applications in the field of X. The viewpoint will be practical throughout. Second semester, MWF at 10.

"Instructor to be announced."

The problem now arises of finding someone to teach the new course. The few well-known statisticians in the country have positions elsewhere from which it would be impossible to dislodge them with the bait to be offered; for though the department wishes to have statistics taught as an auxiliary to the study of X, it feels that there must be no question of the tail wagging the dog, and that economy is appropriate in this connection. The members of the department of professorial rank do not respond favorably to the suggestion that they should themselves undertake to teach the new and unfamiliar course. But every university department has a bright graduate student whose placement is an immediate problem. Young Jones has already demonstrated a quantitative turn of mind in the course on Money and Banking, or in the Ph.D. thesis on which

he has already made substantial progress, dealing with *The Proportion of Public School Yard Areas Surfaced with Gravel*. He may even recall having had a high-school course in trigonometry. His personality is all that might be desired. He is a white, Protestant, native-born American. And so the "Instructor to be announced" materializes as Jones.

This earnest young scholar now finds that, in addition to completing his thesis, he must look up the literature of statistics and prepare a course in the subject. His attention is directed by older members of the department to some of the research papers in the field of *X* involving statistics. He pursues "statistics" through the library card catalog and the encyclopedias. He reads about census and vital statistics, price statistics, statistical mechanics. Perhaps he encounters probable errors. Eventually he learns that Karl Pearson is the great man of statistics, and that *Biometrika* is the central source of information. Unfortunately most of the papers in *Biometrika* and of Pearson's writings, while not lacking in vigor, trail off into mathematical discourse of a kind with which young Jones feels ill at ease. What he wants is a textbook, couched in simple language and omitting all mathematics, to make the subject clear to a beginner. Perhaps he finds the impressive books of Yule and Bowley, but decides that they are too abstruse. Elderton's "Frequency Curves and Correlation" is far too mathematical. Jones decides that a simple book on statistics must be written, and that he will do it if he can ever succeed in mastering the subject. In the meantime, he contents himself perforce with the less mathematical writings of Karl Pearson, with applied examples in the field of *X*, and with such nonmathematical textbooks as may have been written by other young men who have earlier trod the same path as that on which Jones is now beginning. Somehow or other he gets the class through the course. After doing this two or three times, Jones is an experienced teacher of statistics, and his services are much in demand. His course expands, takes on a settled form, and after a while crystallizes into a textbook. At the same time he may be getting out some research, consisting of studies in the field of *X* in which statistical methods play a part. His promotion is rapid. He becomes a Professor of Statistics, and perhaps an officer in a national association. His textbook has a large sale, and is used as a source by other young men writing textbooks on statistics.

The textbooks written in this way form an interesting literary cycle. Measures of "central tendency" and of dispersion are introduced, and the use of one as against another of these measures is debated on every ground except the criterion that modern research has shown to be the important one, the sampling stability. Sampling considerations, indeed, get little attention. The urge to simplify by leaving out the more difficult parts of the subject, and especially the mathematical parts, is accompanied by pride in the great number of examples drawn from real life, that is, actual data that have been collected.

But the most fascinating feature of this literary cycle is the opportunity it offers for research by the standard methods of literary investigation, tracing the

influence of one author upon another through parallelism of passages, and so forth. This study is facilitated by the accumulation of errors with repeated copying. One outstanding example is in certain formulae connected with the rank correlation coefficient, derived originally by Karl Pearson in 1907 and copied from textbook to textbook without adequate checking back. As one error after another was introduced in this process, the formulae presented to students (and apparently made the basis of class exercises involving numerical substitution) became less and less like Pearson's original equations. Incidentally, in trying to check this original work of Pearson's, recent investigation has raised the suspicion that it is erroneous; at any rate, he does not give a fully adequate argument. Thus it may be that the errors in copying, which are so useful in examining the history of statistics, never did any harm. The formulae in which the students were drilled may have been no worse than they would have been if all the copying had been done with more care.

While this process has been going on in the Department of X, the Y and Z Departments have likewise evolved the teaching of statistics. There is some interchange of ideas between the various statisticians on the campus, and there is a catholicity in the copying of textbooks. But by and large, statistics is regarded in the Economics Department as a branch of economics, in the Psychology Department as a part of psychology, and so forth. The astronomer is inclined to resent the suggestion that his students should be called upon to study their least squares with anyone but an astronomer. Medical and biological investigators suspect Economics and Psychology of charlatanry, and do not look with favor on the idea of turning their own students over to such departments for instruction in statistics. Most unthinkable of all would be putting the Department of Education in charge of an essential part of the training of scientific students. Thus the courses multiply.

The fact that it is essentially the same fundamental subject that is being taught under various names and with various kinds of notation in different departments is often concealed by including the teaching of statistical theory in a course whose title and prospectus are more suggestive of applications. A case in point is that of an economist of my acquaintance, not primarily engaged in teaching, who some years ago was invited to give a course in Price Forecasting in the Economics Department of a leading university. He carefully prepared a series of lectures on this subject, which had been the center of some extended research he had conducted. A large class enrolled for the course. But soon after beginning his series of lectures the economist noticed that the class was growing restive. Upon inquiring what was amiss, he learned that his discourse was unintelligible to many of them because he was using technical statistical terms and concepts with which they were not familiar. He thereupon undertook to use simpler language, and when this did not suffice to convey his meaning, to explain the statistical notions involved in his work on price forecasting. More and more his lectures came to deal with the elements of statistics, and less and less with price forecasting. At the end of the term he felt that he had

given the students some elementary knowledge of statistical theory, for which they had not enrolled and for which he did not feel particularly well qualified, but had taught them virtually nothing about price forecasting. When the invitation was repeated the next year, the economist suggested imposing a course in statistics as a prerequisite for the course in Price Forecasting. This however was vetoed by the head of the Economics Department, who did not believe in prerequisites. The Price Forecasting course was not repeated.

This incident illustrates the evolution of a good deal of statistical teaching. At the beginning, the idea is to teach some application, but the teacher soon finds himself engaged at much more length than expected with the fundamentals of statistical theory and methods. In this way it has come about that a large number of persons are teaching theoretical statistics who initially had no intention of doing so, but were concerned with particular applications. The teaching of statistical theory has been undertaken belatedly and inexpertly because it was necessary to a discussion of some application originally in view. Thus it happens that a good deal of teaching of statistics, even of mathematical statistics, masquerades as something else.

The obvious inefficiency of overlapping and duplicating courses given independently in numerous departments by persons who are not really specialists in the subject leads to the suggestion that the whole matter be taken over by the Department of Mathematics. This is a promising solution, but it is doomed to failure if, as has sometimes happened, it means that the teaching of statistics is put under the jurisdiction of those who have no real interest in it. Moreover the teaching of statistics cannot be done appreciably better by mathematicians ignorant of the subject than by psychologists or agricultural experimenters ignorant of the subject. The latter indeed have a certain advantage in that the problems seem more real and definite to them; they can sense the difference between the important and the unimportant questions, even if they cannot express the questions in clear mathematical language, and can sometimes arrive intuitively at a correct result that leaves the mathematician puzzled. Also, they can understand more readily than can the mathematician the examples, drawn largely from biological material, which play so important a part in some of the leading expository work on statistics, such as R. A. Fisher's *Statistical Methods for Research Workers*. The pure mathematician has only one advantage over the non-mathematical worker in empirical fields: he is able to set about reading the serious literature of statistical theory. But he must still find this scattered literature, sort it out from a mass of rubbish, fallacies, and false starts, and trace it back historically until he can understand the notation and the pre-suppositions. He must also contend with the fact that a good deal that is important in statistics is still a matter of oral tradition, and some consists of laboratory techniques. In short, he needs a teacher before he himself sets out to teach the subject. When a Department of Mathematics calls in a young Ph.D., however brilliant, to teach statistics as a part or all of his program, the best thing it can do, if he has not already had a training in modern statistics, is to

give him a furlough for a year or two to enable him to go where he can acquire such a training.

Qualifications of a good teacher of statistics include, first and foremost, a thorough knowledge of the subject. This statement seems trivial, but it has been ignored in such a way as to bring about the present unfortunate situation. Mathematicians and others, who deplore the tendency of Schools of Education to turn loose on the world teachers who have not specialized in the subjects they are to teach, would do well to consider their own tendency to entrust the teaching of statistics to persons who not only have not specialized in the subject, but have no sound knowledge of it whatever. A knowledge of theoretical statistics is not easy to obtain. There is no comprehensive treatise on the subject, starting from first principles, and proceeding by sound deductions and well-chosen definitions to the methods that need to be used in practice. (I have been trying for years to write such a treatise, but it has turned out to be a bigger task than at first appeared. This is partly because some things formerly thought to have been proved turn out, on critical examination, not to be sound, and much new research has been necessary.) The literature is scattered through journals pertaining primarily to many kinds of applications, and it is only in recent years that any large proportion of the current contributions to statistical theory and methods have been gathered into a few periodicals devoted to statistical theory. On the other hand, the seeker after truth regarding statistical theory must make his way through or around an enormous amount of trash and downright error. The great accumulation of published writings on statistical theory and methods by authors who have not sufficiently studied the subject is even more dangerous than the classroom teaching by the same people.

A good teacher of statistics needs of course a mathematical background, including at least an acquaintance with the theory of functions and  $n$ -dimensional euclidean geometry. A good deal of additional algebra and analysis are likely to be helpful, as well as some differential geometry. But no amount of such mathematics constitutes by itself any approach to sufficiency in the qualifications of a teacher of statistics. The most essential thing is that the man shall know the theory of statistics itself thoroughly from the ground up, including the mathematical derivations of proper methods and a clear knowledge of how to apply them in various empirical fields. In addition to the pure mathematics and the knowledge of statistical theory, a competent statistician or teacher of statistics needs a really intimate acquaintance with the problems of one or more empirical subjects in which statistical methods are applied. This is quite important. Sometimes excellent mathematicians have wasted time and misled students through failure to get that feeling for applications that is necessary for proper statistical work.

The theory of statistics has been making advances so rapid and so fundamental that some of the first things that need to be said in an elementary course, even for prospective practical statisticians, are affected by some of the most recent researches. So elementary a question as "What definition is it wise to give to

the term 'standard deviation'?", which must be faced by every teacher of Statistics 1, requires for an intelligent answer a rather thorough understanding of modern sampling theory and techniques. The answer, it now seems, is *not* the definition given in most textbooks. In the selection of a statistic to represent a parameter, for example in fitting frequency curves or in linkage estimation in genetics, the fundamental consideration is connected with the sampling distribution, as R. A. Fisher showed in founding the modern theory of estimation. This is ignored in most of the current teaching of statistics, with the result that innumerable students are sent out to waste the money and time of their employers by demanding larger samples than are necessary for the purposes in view, wasting costly information by calculating inefficient statistics and using tests that are not the most powerful. On the other hand, students of statistics who are taught rule-of-thumb methods without their derivations are never quite conscious of the exact limitations and assumptions involved, and may make unwarranted inferences from samples that are too small or in some way violate the conditions underlying the derivations of the formulae.

A good teacher of statistics must be thoroughly familiar with these recent advances. He must examine very critically textbook statements unsupported by full proofs. Even though the students are not capable of following the complete mathematical argument—indeed, especially if the students are not to examine it—the instructor needs to give it a critical study. The custom of omitting proofs, which would not be tolerated in pure mathematics beyond a very limited extent, is common in the teaching of statistics, and is excused on the ground that the students do not know enough mathematics to understand the proofs. Perhaps in some cases a better reason is that the teachers, and the authors of the textbooks, do not understand the proofs. In some instances no proofs exist, and in some instances no genuine proofs can exist, because the methods taught are demonstrably wrong. The custom prevalent in the teaching of mathematics of going over each proof carefully in the class is, among other things, a safeguard against infiltration of false propositions. This safeguard is missing from most of the teaching of statistics, and there has been an infiltration of errors. Since it is accepted that a great many students need to learn something about statistical methods without learning enough mathematics to understand the proofs, it follows that the elementary teaching of statistics to these students must, if the perpetuation of gross errors is to be avoided, be in the hands of really competent mathematical statisticians. This is perhaps the greatest reform needed in the teaching of statistics today. Until the *elementary* teaching of statistics is conducted by those with a thorough and critical knowledge of current research in statistical theory, of a sort that seems virtually inseparable from participation in that research, there is likely to be a continuation of the laborious drilling of thousands of students in methods that ought never to be used. Here, of all places, is the great need for participation of research workers in elementary teaching.

Teachers and textbook writers might well abandon the idea of telling what

statistical methods are used, and say instead what methods ought to be used. But before they can do this with confidence they must have a very close acquaintance with the research of the last three decades in statistical theory.

How can an appointing officer know whether a prospective teacher of statistics knows his subject? This question requires no answer peculiar to statistics in distinction from other subjects. Publication of research, constituting a contribution to the particular field, has always been accepted as the best proof. A substantial contribution to fundamental statistical theory, which is to be distinguished from the mere application of known statistical methods to empirical data, is the best indication of the kind of scholarship appropriate to a teacher of statistics.

Participation in research is not novel as a criterion of what constitutes a good teacher of a college or university subject, if the subject is Greek literature, physics, chemistry, biology, or indeed any of those departments that have been long enough established to attain with respect to the organization of their teaching a state approximating equilibrium. The more reputable institutions of higher learning have long maintained the principle, though with occasional violations in practice, that the Ph.D. degree or its equivalent, representing among other things the completion of a piece of scholarly research, is a minimum condition for a regular faculty appointment. It has usually been maintained also that the Ph.D. thesis should be a new contribution of a strictly scholarly character to the field of the scholar's competence, and not merely a routine application of known methods to an extraneous field. Thus a thesis offered for the Ph.D. degree in mathematics would be judged by its contribution to mathematics, rather than to physics or accounting. Moreover the regard in which universities have held members of their faculties has been intimately connected with their output of scholarly research. Other criteria of excellence have not been ignored, but research has been recognized in a fairly consistent manner. Some say that there has been an over-emphasis on research, and that more attention ought to be given to other qualities related to teaching. However this may be, the facts remain that scholarly research is something capable of a reasonably objective evaluation by scholars in the field, that it offers the main hope of fundamental progress, and that familiarity with current research is a necessary, though not sufficient, condition for the most important teaching in institutions of higher learning.

A peculiarity of the teaching of statistics, of which in practice the theory of statistics is an essential even if unacknowledged part, is that a good deal of it has been conducted by persons engaged in research, not of a kind contributing to statistical theory, but consisting of the application of statistical methods and theory to something else. A similar situation would exist if the teaching of mathematics were in the hands of an assortment of various kinds of engineers, or if zoology and botany were taught by practicing physicians. The teaching of mathematics and of elementary biology might perhaps gain in liveliness and concreteness by such arrangements, with the accompanying emphasis on the



particular applications of the fundamental sciences. Moreover the engineer might in the course of such teaching refresh his own knowledge of elementary mathematics, while the physician might gain by renewing his acquaintance with elementary biology. Such arrangements might occasionally be made with profit. But if they were the general rule the advantages of specialization would be lost; the fundamental sciences would not be developed in so well-rounded a manner as they are by specialists in them, while the special skills and knowledge of the physician and engineer could not be utilized to the full in their respective professions. Statistical theory is a big enough thing in itself to absorb the full-time attention of a specialist teaching it, without his going out into applications too freely. Some attention to applications is indeed valuable, and perhaps even indispensable as a stage in the training of a teacher of statistics and as a continuing interest. But particular applications should not dominate the teaching of the fundamental science, any more than particular diseases should dominate the teaching of anatomy and bacteriology to pre-medical students. These subjects are not ordinarily taught by practicing physicians, but by anatomists and bacteriologists respectively.

In medical education the principle has been accepted, after a long struggle, that a medical school should have full-time professors engaged primarily in teaching and research, and that such professors should not treat patients except in cases of unusual interest from the standpoint of the science or art of medicine. An analogous principle would be that an institution offering extensive instruction in statistics should have full-time professors engaged in the teaching of and research in statistical theory and methods, without spending time over applied statistical problems excepting insofar as such problems might present novel features calling for the development of new statistical methods or theoretical extensions having interest going beyond the immediate case. Sometimes the complaint is heard in medical schools that the teaching tends to become too theoretical on account of detachment from clinical practice, and a similar difficulty might conceivably develop in connection with statistics; but in neither case does the trouble seem to be beyond the ability of the personnel involved to cure if they have the right background.

A specialist in statistics on a university faculty has a threefold function. In addition to the usual duties of teaching and research, there is a need for him to advise his colleagues, and other research workers, regarding the statistical methods appropriate to their various investigations. The advisory function is a highly important one for the activities of the university as a whole, and should be taken into consideration in adjusting the teaching load. Probably every university statistician is visited from time to time by earnest research workers, deeply engrossed in their respective specialties, speaking technical jargons unfamiliar to the statistician, and seeking his advice on matters concerning which he has a sinking feeling of lack of comprehension. After some hours of psychoanalyzing his visitor the statistician may be able to ascertain what it is he *really* wants to know, and thereafter either refer him to some standard formula, or

more often, undertake a piece of new mathematical research designed to fit the particular problem, and very possibly having value also for a more extended class of problems. The statistician is then very likely to find himself embarked on a co-operative research venture in a field that is new to him.

To function well in this third, the consultative or co-operative function, he must have an unusually large store of general information. No one stands in greater need than he of that knowledge of "something about everything and everything about something" that was once said to be the goal of a liberal education. In planning the education of statisticians and teachers of statistics these considerations point to a somewhat wider diffusion of studies among various fields than is customary in many institutions, especially in graduate work. The co-operation, and their other work, would also be facilitated if research workers in general were more strongly urged to get a training in mathematical statistics at an early stage in their careers.

The problem of departmental organization is secondary to that of getting men having the requisite qualities of extensive mathematical preparation, a thorough knowledge of modern theoretical statistics, an understanding of some fields at least in which statistical methods can be applied, and the type of inquiring mind sometimes described as a "research outlook." A Department of Mathematics may well handle the fundamental teaching in statistics, provided it has men properly qualified for such teaching. If it does not have such men, its teaching of statistics and its inability to provide the needed statistical advice will inevitably tempt the other departments to set up again their own duplicating courses in what amounts essentially to statistical theory and methods, and to repeat the mistakes of the past.

A separate Department of Statistics, if competently staffed, could very well provide advice for the whole institution as well as conducting elementary instruction in statistical methods and theory, both for students having calculus and for those without it, and should certainly carry on advanced teaching and research in statistical theory and methods. But for efficient functioning of the institution as a whole it should be agreed that the Department of Statistics or the Department of Mathematics should do *all* the elementary instruction in statistics, and that courses in statistics in other departments should be confined to applications of the basic theory. Normally such courses in applied statistics in the other departments should require as a prerequisite one or more of the basic courses in the Department of Statistics, or of Mathematics. The basic course to be required as a prerequisite to others should be the one which itself requires calculus as a prerequisite wherever this is practicable. It is practicable for students of engineering, physics, astronomy, and mathematical economics, since these students must have calculus anyhow. Moreover the value of the sequence consisting of calculus, statistical theory and applied statistics, in this order, is so great that many other students are likely to avail themselves of it when it is once established and the true nature and value of statistics are more widely understood.

Exactly how far a Department of Statistics should go in particular applications would have to be decided anew from time to time by its members in the light of changing conditions and interests. It cannot teach everything that goes by the name of statistics. This problem may be exemplified by the case of population and vital statistics. This is a field with close connections with sociology, biology, medicine and insurance. It is cultivated in conjunction with each of these subjects in various places. Some of its most interesting and important phases make use of quite advanced mathematics, as in the work of A. J. Lotka, and in addition there is extensive use, and more extensive need, of the statistical methods centered around sampling theory which are the appropriate domain of a Department of Statistics. Should the study of population and vital statistics be included in a Department of Statistics? I think not, except as a temporary arrangement, or in a small institution, in spite of the history of the word "statistics," which originated in connection with material of this kind, and in one of its meanings is still applied to it. (My use of the unqualified word "statistics" in this paper is in the sense of theory and methods, not in the sense of statistical facts such as those found by the census.) Medical, biological and sociological considerations are prominent in the problems of vital statistics, and one of these departments might well handle the subject. But the vital statistician, like other research workers, should have acquired in the course of his training an intimate familiarity with the statistical theory and methods which are the appropriate province of a Department of Statistics. He also needs mathematics through integral equations, if he is to understand and extend the contributions of Lotka and Volterra. Students of vital statistics should have had an elementary course in statistical theory in the Department of Statistics, preferably the course requiring calculus.

A course in price statistics should be taught by an economist, presumably in the Department of Economics, but might well require as a prerequisite the same elementary courses in statistical theory and methods as would be required in psychology, medicine and other fields. In addition, there are problems of time series analysis whose treatment calls for a mathematical statistician having some acquaintance with both economic and meteorological data. A course on the treatment of time series might appropriately be included in the Department of Statistics, requiring the general elementary course as a prerequisite, and itself serving as a prerequisite for courses in economic and meteorological statistics.

One of the chief obstacles to efficient organization of teaching is the habit of not prescribing prerequisites outside one's own department. But when once the elementary courses in statistics have become established in the hands of well-equipped specialists in statistical theory and methods, in whose competence general confidence can be reposed, the various departments of application will lose their motive for establishing their own duplicating courses, and will be able to cultivate more intensively their respective specialities.

The detection of biases and the details of practical statistical work vary greatly

from one application to another. These, consequently, are matters for the departments concerned with applications rather than with the fundamentals of statistics, and should not be the chief features of a course in elementary statistical methods and theory. The work of a Department of Statistics should be concerned largely with sampling theory, and should emphasize the unity of statistical methods and theory, regardless of the field of application. It should deal with statistics as a coherent science of inductive inference, of the preparation of observations for inference, and of the planning of investigations so as to yield observations from which inferences can best be made.

The question what mathematical prerequisites should be established for the fundamental course in statistical theory must be answered by a compromise between the ideal and what is expedient at a particular time and place. In Europe a large number of students have had a year of calculus before coming to universities, that is, before reaching the age of eighteen. If a university were willing to restrict its entrants to such students (thus automatically solving the problem of overcrowding) it could give them another year of calculus, mixed perhaps with advanced algebra and geometry, and then in their sophomore year give them a thorough course in elementary statistics and probability, based on calculus. These students would then be ready to tackle advanced statistics in the third year in a really effective way. If the teaching of economic theory, physics, chemistry and astronomy were geared to this program in such a way as to make real use of the calculus, the work in these subjects could be made far more efficient, in the sense that more material could be covered effectively in the allotted time, or an equivalent amount of material in less time. If, in addition, all the many departments in which statistical methods and theory are used required these statistical courses as prerequisites, and actually used the materials of these courses in their work, there would be a further huge gain in efficiency. The baccalaureate degree of such an institution would represent a far more thorough knowledge, and command of the tools of research, than is possible without an arrangement putting in this way the fundamentals first.

Institutions unwilling to undertake such a drastic improvement must face more or less delay and inadequacy in the acquisition by their students of the fundamentals of mathematics and of statistics. A division of the students into groups according to mathematical ability ought to be undertaken, and followed by a corresponding division of the elementary statistics course. Students having high mathematical ability could begin the study of statistics after completing calculus, and could look forward to rising ultimately to greater heights in pursuits involving mathematical or statistical knowledge than those of lesser mathematical talents. For these latter there would still be the possibility of acquiring, even without calculus, useful statistical tools; but it is essential that this should be done under the guidance of instructors thoroughly familiar with the mathematics of statistics. The task of leading the blind must not be turned over to the blind. Students possessing the ability to master the calculus should

be encouraged to begin the study of statistics with the course having calculus as a prerequisite, and should not be put into the necessarily slower group not having the calculus. I believe that these elementary courses should begin with the theory of probability, but should go on to the chief distribution functions used in practice, and should include applied problems and work on calculating machines.

Putting a sound program of statistical teaching into effect will take time, partly because of the scarcity of suitable teachers of statistics. Nevertheless, the process is well under way, and the prospects are good for substantial improvements in the teaching of statistics. A body of able young research men possessing the requisite knowledge of statistical fundamentals is now in existence and is growing. Some of the recent textbooks represent striking improvements. The Institute of Mathematical Statistics itself, with the *Annals of Mathematical Statistics*, is perhaps the best evidence of a changed view making for better things.

COLUMBIA UNIVERSITY,  
NEW YORK, N. Y.

## DISCUSSION OF PROFESSOR HOTELLING'S PAPER

By W. EDWARDS DEMING

It is a pleasure to endorse Professor Hotelling's recommendations; in fact we have been following them pretty closely in the courses in the Graduate School of the Department of Agriculture. As a matter of fact, he has indirectly played an influential part in building up this set of courses, because some of our best instructors are his former students.

Listening to Professor Hotelling's paper, I was thinking of the possibility that some of his recommendations might be misunderstood. I take it that they are not supposed to embody all that there is in the teaching of statistics, because there are many other neglected phases that ought to be stressed. In the Bureau of the Census the population division alone has augmented its force by approximately 3500 statistical clerks during the past six months. They come from diverse schools and it has been interesting to observe how many of them have the idea that all the problems of sampling and inference from data can be solved by what are commonly known as modern statistical techniques—correlation coefficients, rank correlation coefficients, chi-square, analysis of variance, confidence limits, and the like. Most of them are shocked to learn that many of the so-called modern "theories of estimation" are not theories of estimation at all, but are rather theories of distribution and are a disappointment to one who is faced with the necessity of making a prediction from his data, i.e., of basing

some critical course of action on them. The conviction that such devices as confidence limits and Student's  $t$  provide a basis for action regardless of the size of the sample whence they were computed, even under conditions of statistical control, is too common a fallacy. On the other hand, many simple but worthy devices are neglected. A histogram, for instance, can be a genuine tool of prediction if it is built up layer by layer in different legends so as to distinguish the different sources whence the data are derived. The modern student, and too often his teacher, overlook the fact that such a simple thing as a scatter diagram is a more important tool of prediction than the correlation coefficient, especially if the points are labeled so as to distinguish the different sources of the data. Most students do not realize that for purposes of prediction the consistency or lack of it between many small samples may be much more valuable than any probability calculations that can be made from them or from the entire lot. Students are not usually admonished against grouping data from heterogeneous sources. Of those that are not guilty of indiscriminate grouping, many are inclined to rely on statistical tests for distinguishing heterogeneity, rather than on a careful consideration of the sources of the data. Too little attention is given to the need for statistical control, or to put it more pertinently, since statistical control (randomness) is so rarely found, too little attention is given to the interpretation of data that arise from conditions not in statistical control.

Nevertheless, the fundamentals of probability and sampling theory, and the mathematics of the distribution functions, though by themselves they do not qualify anyone for high-grade statistical work, are ultimately essential for proficiency in statistics. Since they are seldom learned away from the university they are properly made the main theme of teaching. The university is the place to learn the studies that are so difficult to get outside of it.

Above all, a statistician must be a scientist. The skepticism of many first class scientists of today for modern statistical methods should be a challenge to statistical teaching. A scientist does not neglect any pertinent information, yet students of statistics are often taught to do just the opposite of this, and are accused of being old-fashioned for daring to think of combining experience with the new information provided by a sample, even if it is a pitifully small one. Statisticians must be trained to do more than to feed numbers into the mill and grind out probabilities; they must look carefully at the data, and take account of the conditions under which each observation arises. It is my feeling that the chief duty of a statistician is to help design experiments in such a way that they provide the maximum knowledge for purposes of prediction; another is to compile data with the same object in view; and still a third function is to help bring about some changes in the source of the data. Scientific data are not taken merely for inventory purposes. There is no use taking data if you don't intend to do something about the sources whence they arise.

## RESOLUTIONS ON THE TEACHING OF STATISTICS

The Institute of Mathematical Statistics at its business meeting on September 11, 1940 at Dartmouth College adopted the following resolutions regarding the teaching of statistics. The resolutions were drawn up by a committee appointed by the President, and consisting of Burton H. Camp, W. Edwards Deming, Harold Hotelling, and Jerzy Neyman.

1. If the teaching of statistical theory and methods is to be satisfactory, it should be in the hands of persons who have made comprehensive studies of the mathematical theory of statistics, and who have been in active contact with applications in one or more fields.

2. The judgment of the adequacy of a teacher's knowledge of statistical theory must rest initially on his published contributions to statistical theory, in contrast with mere applications, in a manner analogous to that long accepted in other university subjects.

3. These ideas are expressed in detail in the paper *The teaching of statistics*, by Professor Harold Hotelling, and the Institute decides to give both the resolution and the paper as wide a circulation as possible.

## REPORT OF THE HANOVER MEETING OF THE INSTITUTE

The sixth meeting of the Institute of Mathematical Statistics was held at Dartmouth College, Hanover, New Hampshire, Tuesday to Thursday, September 10 to 12, 1940, in conjunction with meetings of the American Mathematical Society and of the Mathematical Association of America. The following forty-two members of the Institute attended the meeting:

H. E. Arnold, Felix Bernstein, G. W. Brown, J. H. Bushey, B. H. Camp, A. T. Craig, A. R. Crathorne, J. H. Curtiss, J. F. Daly, W. E. Deming, J. L. Doob, Churchill Eisenhart, M. L. Elveback, C. H. Fischer, M. M. Flood, R. M. Foster, T. C. Fry, H. P. Geiringer, Robert Henderson, E. H. C. Hildebrandt, G. M. Hopper, Harold Hotelling, E. V. Huntington, M. H. Ingraham, Dunham Jackson, W. L. Kichline, L. F. Knudsen, B. A. Lengyel, W. G. Madow, J. W. Mauchly, Richard von Mises, E. B. Mode, Jerzy Neyman, P. S. Olmstead, Oystein Ore, M. M. Sandomire, L. W. Shaw, F. F. Stephan, A. G. Swanson, Abraham Wald, S. S. Wilks, Jacob Wolfowitz.

The meeting of the Institute consisted of four sessions. At the first session, which was held on Tuesday morning, Professor Harold Hotelling of Columbia University delivered an address on *The Teaching of Statistics*. This address was followed by considerable discussion on the various aspects of the teaching of statistics.<sup>1</sup> Preceding Professor Hotelling's address a short paper on an *Empirical Comparison of the "Smooth" test for goodness of fit with Pearson's Chi-Square test* was presented by Professor J. Neyman of the University of California.

Following Professor Hotelling's address a business meeting of the Institute was held. At this time resolutions on the teaching of statistics were approved (see p. 472). The President reported that a War Preparedness Committee had been appointed in the summer to study the matter of the Institute's participation in the national defense program.<sup>2</sup> The Chairman of this Committee submitted a preliminary report which met the approval of the Institute. A plan was approved for completing the report and circularizing it with a minimum of delay.

The matter of the organization of local sections or chapters of the Institute was discussed but no action was taken.

<sup>1</sup> Professor Hotelling's address and three resolutions regarding the teaching of Statistics which were adopted by the Institute at a business meeting following the address are published in the present issue of the *Annals of Mathematical Statistics*, pp. 457-472.

<sup>2</sup> The membership of the Committee is as follows:

Professor Churchill Eisenhart (Chairman), University of Wisconsin.

Professor A. T. Craig, University of Iowa.

Professor E. G. Olds, Carnegie Institute of Technology.

Captain Leslie E. Simon, Aberdeen Proving Ground.

Mr. Ralph E. Wareham, General Electric Company.



On Tuesday afternoon a session on contributed papers in Mathematical Statistics was held jointly with the American Mathematical Society. Professor B. H. Camp of Wesleyan University presided and the following papers were presented:

1. *Contributions to the theory of the representative method of sampling.*  
Dr. W. G. Madow, Department of Agriculture, Washington.
2. *A generalization of the law of large numbers.*  
Dr. Hilda P. Geiringer, Bryn Mawr College.
3. *On the problem of two samples from normal populations with unequal variances.*  
Professor S. S. Wilks, Princeton University.
4. *Experimental determination of the maximum of an empirical function.*  
Professor Harold Hotelling, Columbia University.
5. *Asymptotically shortest confidence intervals.*  
Dr. Abraham Wald, Columbia University.
6. *Reduction of certain composite statistical hypotheses.*  
Dr. G. W. Brown, R. H. Macy and Company, Inc., New York.
7. *Conception of equivalence in the limit of tests and its application to certain  $\lambda$  and  $\chi^2$  tests.*  
Professor J. Neyman, University of California.

Abstracts of these papers follow this report.

On Wednesday morning a session was held on *The Theory of Probability* with Dr. T. C. Fry of the Bell Telephone Laboratories, in the chair. The following addresses were given:

1. *On the foundations of probability theory.*  
Professor R. von Mises, Harvard University.
2. *Probability as measure.*  
Professor J. L. Doob, University of Illinois.

This session was followed by an energetic discussion which was continued in an informal afternoon session.

The Thursday morning session was devoted to the *Theory of Statistical Estimation* with Professor Harold Hotelling as Chairman. The following addresses were given:

1. *Estimation by intervals as a classical problem in probability.*  
Professor J. Neyman, The University of California.
2. *Statistical estimation in large samples.* Dr. Joseph F. Daly, The Catholic University of America.

On Monday at 4:15 p.m. a tea was held at the Graduate Club for members of the mathematical organizations and their guests, and on Monday at 8:00 a musical performance was presented. On Tuesday at 7:00 p.m. a joint dinner was held for the mathematical organizations in Thayer Hall. Wednesday afternoon was devoted to an excursion to Franconia Notch.

During the meeting a collection of string models of ruled surfaces was exhibited by Professor Robin Robinson of Dartmouth College and electrical calculation apparatus made from telephone equipment was exhibited by members of the staff of the Bell Telephone Laboratories.

## ABSTRACTS OF PAPERS

(Presented on September 10, 1940, at the Hanover meeting of the Institute)

### Contributions to the Theory of the Representative Method of Sampling. WILLIAM G. MADOW, Washington, D. C.

The theory of representative sampling may be regarded as a dual sampling process; the first of which consists in the sampling of different random variables and the second of which consists in repeating several times the experiments associated with each of the different random variables. It follows that while the theory of sampling from finite populations without replacement may be required for the first process, the second leads directly into the theory of sampling from infinite populations. There is, however, one difference. Although the usual theory is concerned with the evaluation of fiducial or confidence limits for parameters the theory of sampling is concerned with the evaluation of fiducial or confidence limits for, say, the mean of a sample of  $N$ , when  $n$ , ( $N \geq n$ ), of the values are known.

It is thus possible to use the usual theories of estimation in obtaining estimates of the parameters and to allow the effects of subsampling process to show themselves in the different values of the fiducial limits. It is shown that the limits obtained are almost identical with those obtained by the theory of sampling from a finite population. Distributions of the statistics used in these limits are derived.

Besides these results, the theory is extended to the theory of sampling vectors, and conditions are stated under which the "best" allocation of the number in a sample among several strata is proportional to the  $k$ th roots of the generalized variance of a random vector having  $k$  components.

### A Generalization of the Law of Large Numbers. HILDA GEIRINGER, Bryn Mawr.

Let  $V_1(x)$ ,  $V_2(x)$ ,  $\dots$ ,  $V_n(x)$  be  $n$  probability distributions which are not supposed to be independent and let  $F(x_1, x_2, \dots, x_n)$  be a "statistical function" of  $n$  observations in the sense of v. Mises,— $V_i(x)$  ( $i = 1, 2, \dots, n$ ) indicating as usual the probability of getting a result  $\leq x$  at the  $i$ th observation—. Then it can be proved that under fairly general conditions  $F(x_1, x_2, \dots, x_n)$  converges stochastically toward its "theoretical value"; or in other words, that under these general conditions a great class of statistics  $F(x_1, x_2, \dots, x_n)$  is "consistent" in the sense of R. A. Fisher.

Well known particular cases of this theorem result if (a) we take for  $F(x_1, x_2, \dots, x_n)$  the average  $(x_1 + x_2 + \dots + x_n)/n$  of the  $n$  observations, (b) we assume that the  $V_i(x)$  are independent distributions.

### On the Problem of Two Samples from Normal Populations with Unequal Variances. S. S. WILKS, Princeton University.

Suppose  $O_{n_1}$  and  $O_{n_2}$  are samples of  $n_1$  and  $n_2$  elements from normal populations  $\pi_1$  and  $\pi_2$  respectively. Let  $a_1$ ,  $\sigma_1^2$  and  $a_2$ ,  $\sigma_2^2$  be the means and variances of  $\pi_1$  and  $\pi_2$  and let  $O_{n_1}$  and  $O_{n_2}$  have means  $\bar{x}_1$  and  $\bar{x}_2$  and variances  $s_1^2$  and  $s_2^2$  (unbiased estimates of  $\sigma_1^2$ ,  $\sigma_2^2$ ) respectively. It is shown that there exists no function (Borel measurable) of  $\bar{x}_1$ ,  $\bar{x}_2$ ,  $s_1^2$ ,  $s_2^2$ ,  $a_1 - a_2$  independent of  $\sigma_1$  and  $\sigma_2$ , having its probability law independent of the four population parameters. It is therefore impossible to obtain exact confidence limits

for  $a_1 - a_2$  corresponding to a given confidence coefficient. Functions of the four parameters and four statistics are devised from which one can set up confidence limits for  $a_1 - a_2$  with associated confidence coefficient inequalities.

### Experimental Determination of the Maximum of an Empirical Function.

HAROLD HOTELLING, Columbia University.

In physical and economic experimentation to determine the maximum of an unknown function, for example of a monopolist's profit as a function of price, or of the magnetic permeability of an alloy as a function of its composition, the characteristic procedure is to perform experiments with chosen values of the argument  $x$ , each of which then yields an observation, subject to error, on the corresponding functional value  $y = f(x)$ . The values of  $x$  need, however, to be chosen on the basis of earlier experiments in order to make the determination efficient. The experimentation properly proceeds, therefore, in successive stages, with the values used at each stage determined with the help of the earlier work. The question what distribution of  $x$  as a function of previous results should be used is discussed in this paper on the basis of various hypotheses regarding the function, and further criteria. In particular, a conflict is shown to exist under some conditions between the criterion of minimum sampling variance and that calling for absence of bias.

### Asymptotically Shortest Confidence Intervals. ABRAHAM WALD, Columbia University.

Let  $f(x, \theta)$  be the probability density function of a variate  $x$  involving an unknown parameter  $\theta$ . Denote by  $x_1, \dots, x_n$   $n$  independent observations on  $x$  and let  $C_n(\theta)$  be a positive function of  $\theta$  such that the probability that  $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \sum_{a=1}^n \log f(x_a, \theta) \leq C_n(\theta)$  is equal to a constant  $\beta$  under the assumption that  $\theta$  is the true value of the parameter.

Denote by  $\theta'(x_1, \dots, x_n)$  the root in  $\theta$  of the equation  $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \sum_a \log f(x_a, \theta) = C_n(\theta)$

and by  $\theta''(x_1, \dots, x_n)$  the root of  $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \sum_a \log f(x_a, \theta) = -C_n(\theta)$ . Under some weak

assumptions on  $f(x, \theta)$  the interval  $\delta_n(x_1, \dots, x_n) = [\theta'(x_1, \dots, x_n), \theta''(x_1, \dots, x_n)]$  is in the limit with  $n \rightarrow \infty$  a shortest unbiased confidence interval<sup>1</sup> of  $\theta$  corresponding to the confidence coefficient  $\beta$ . This confidence interval is identical with that given by S. S. Wilks in his paper "Shortest average confidence intervals from large samples," *The Annals of Mathematical Statistics*, Sept. 1938. Wilks has shown that  $\delta_n(x_1, \dots, x_n)$  is asymptotically shortest in the average compared with all confidence intervals computed on the basis of statistics belonging to a certain class  $C$ . In the present paper it has been proved that the confidence interval in question is asymptotically shortest compared with any arbitrary unbiased confidence interval, without any restriction to a certain class of functions.

### Reduction of Certain Composite Statistical Hypotheses. GEORGE W. BROWN, R. H. Macy and Co., New York.

The results obtained make it possible to reduce a large class of composite statistical hypotheses to equivalent simple hypotheses. The fundamental theorem established states essentially that if two distributions give rise, in sampling, to the same distribution of the

<sup>1</sup> For the definition of a shortest unbiased confidence interval see the paper by J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc.* (1937).

set of differences between observations, then one distribution must be a translation of the other, subject to a condition requiring that the characteristic function of one of the distributions be such that any interior intervals of zeros be not too large. The result is established by means of the functional equation  $\varphi(t_1)\varphi(t_2)\varphi(-t_1 - t_2) = \psi(t_1)\psi(t_2)\psi(-t_1 - t_2)$  relating the characteristic functions. Similar results are obtained for scale, and combination of location and scale, and the corresponding situations in multivariate distributions. This type of uniqueness theorem permits one to reduce a composite hypothesis involving an unknown location parameter (or scale, or both) to an equivalent simple hypothesis.

# Conception of Equivalence in the Limit of Tests and Its Application to Certain $\lambda$ - and $\chi^2$ -Tests. J. NEYMAN, University of California.

Denote by  $E$  a system of observable variables and by  $N$  the number of independent observations of those variables to be used for testing a certain statistical hypothesis  $H$  against a set  $\Omega$  of admissible simple hypotheses  $h$ . Let further  $T_1(N)$  and  $T_2(N)$  be two different tests of  $H$  using the same number  $N$  of observations. Consider the probability  $P_N(h)$  calculated on any admissible simple hypothesis  $h$ , of the two tests, contradicting themselves.

*Definition:* If, whatever be  $h \in \Omega$ , the probability  $P_N(h)$  tends to zero as  $N$  is indefinitely increased, then the two tests are said to be equivalent in the limit.

Consider a number  $s$  of series of independent trials and denote by  $E_{i1}, E_{i2}, \dots, E_{im_i}$  all the  $m_i$  possible and mutually exclusive outcomes of each of the trials forming the  $i$ th series. Let  $p_{ij}$  be the probability of  $E_{ij}$ ,  $n_i$  the total number of trials in the  $i$ th series, and  $n_{ij}$  the number of these which give the outcome  $E_{ij}$ .

Suppose that it is desired to test a composite hypothesis  $H$  concerning all the probabilities  $p_{ij}$  and consisting of the assumption that any one of them is a given linear function of some  $t$  independent parameters  $\theta_k$ , so that

$$(1) \quad p_{ij} = a_{ij0} + a_{ij1}\theta_1 + \dots + a_{ijt}\theta_t$$

where the coefficients  $a_{ijk}$  are known. The main result of the paper is then that the  $\lambda$ -test of the above hypothesis  $H$ , tested against the set  $\Omega$  of alternatives ascribing to the  $p_{ij}$  any non-negative values, is equivalent in the limit to the test consisting of rejecting  $H$  when the minimum of the expression

$$(2) \quad \chi^2 = \sum_{i=1}^s \sum_{j=1}^{m_i} \frac{(n_{ij} - n_i p_{ij})^2}{n_{ij}}$$

calculated with respect to unrestricted variation of the  $\theta$ 's, exceeds the tabled value of  $\chi^2$  corresponding to the chosen level of significance  $\epsilon$  and to the number of degrees of freedom

$$\sum_{i=1}^s m_i - s - t.$$

It will be noticed that the expression (2) differs from the usual  $\chi^2$  in the denominator of each term.

As an example of the application of the test based on (2), consider the case where  $M$  varieties of sugar beet are tested for resistance to a certain disease in an experiment arranged in  $N$  randomized blocks. Denote by  $n$  the number of beets selected at random for inspection from each plot and by  $n_{ij}$  the number of those of the  $i$ th variety from the plot in the  $j$ th block which are found to be infected. Denote further by  $p_{ij}$  the proportion of infected beets of the  $i$ th variety in the plot in the  $j$ th block. The hypothesis that the effects of variety and of block are additive is expressed by  $p_{ij} = p + V_i + B_j$  with  $\sum V_i = \sum B_j = 0$ . To test this hypothesis we may use (2) which in this particular case reduces itself to

$$(3) \quad \chi^2 = \sum_{i=1}^M \sum_{j=1}^N w_{ij}(q_{ij} - p - V_i - B_j)^2$$

with  $w_{ij} = n^2/[n_{ij}(n - n_{ij})]$ ,  $q_{ij} = n_{ij}/n$ . The minimum  $\chi_0^2$  of  $\chi^2$  is found by solving a set of equations which are linear in  $p$ ,  $V_i$ ,  $B_j$  and the comparison of  $\chi_0^2$  with the tabled value corresponding to  $(M - 1)(N - 1)$  degrees of freedom will tell us whether we are likely to be very wrong in assuming additivity or not. In the favorable case we may next proceed similarly to test another hypothesis that there is no differentiation between the varieties, so that  $V_1 = V_2 = \dots = V_M = 0$ .

**Empirical Comparison of the "Smooth" Test for Goodness of Fit with the Pearson's  $\chi^2$  Test.** J. NEYMAN, University of California.

In a previous publication<sup>1</sup> the author has deduced a test for goodness of fit, described as the "smooth test" or the  $\psi^2$  test, applicable to cases where the hypothesis tested  $H$  is simple. The test is so devised as to be particularly sensitive to departures from  $H$  which are "smooth" in the sense explained in detail in the publication quoted. Whether the test so devised does present any advantage over the usual  $\chi^2$  test depends on how frequently we meet, in practice, cases where the hypotheses alternative to the one tested are actually smooth.

The present investigation was undertaken with the object of obtaining some information on this point. For that purpose a number of cases described in the literature where there was a question of testing that some observable variable  $x$  follows some perfectly specified distribution  $p(x)$  were analyzed. Of all such cases, the ones where there were *a priori* theoretical reasons to believe that  $p(x)$  could not possibly represent the true distribution of  $x$  and, at the most, it could be considered as only an approximation to the true distribution were selected.

It was assumed that the departures from the hypothetical distributions are typical of those that may be met in practice when no definite information as to the actual state of affairs is available. The hypothesis of goodness of fit was tested both by means of the  $\chi^2$  and by the fourth order smooth test. Out of the 130 cases studied the two tests were in perfect agreement eight times. Out of the remaining 122 cases the smooth test proved to be more sensitive than the  $\chi^2$  in 70 cases and the  $\chi^2$  better than the smooth test in 52 cases. We may further compare the tests by counting those cases where one of them detected the falsehood of the hypothesis tested at a given level of significance while the other failed to do so. At the level of significance .05 the  $\chi^2$  test rejected the hypothesis tested 13 times, while  $P_{\psi^2}$  was  $>.05$ . The reverse was true in 17 cases. At the level of significance .01 the corresponding figures are 5 and 14, again in favor of the smooth test.

<sup>1</sup> J. Neyman, " 'Smooth Test' for Goodness of Fit." *Skandinavisk Aktuarietidskrift*, 1937, pp. 149-199.

## REPORT OF THE WAR PREPAREDNESS COMMITTEE OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The generally recognized functions of a *statistician* are the calculation of averages, percentages, and index numbers; the construction of bar graphs and pie diagrams; and the compilation of data in general. His other activities are less widely known. In particular, the recent advances in *mathematical statistics* are known to a relatively small proportion of the persons occupying responsible positions in academic life, in industry, and in government. The *mathematical statistician*, in fact, is concerned chiefly with the interpretation of data through the use of probability theory; his is the science of reasoning from a part to the whole, and of prediction; and to him falls the task of stating the conditions under which such inferences are possible, of devising means of testing whether these conditions are satisfied, and of evaluating the probability that such 'uncertain inferences' are correct in specific instances. Furthermore, it is his responsibility to so plan the lay-out of experiments and the conduct of surveys that the data they yield will contain the maximum information on the points at issue and be amenable to unambiguous statistical interpretation.

Because of the functions which the *mathematical statistician* can perform his services should be of value to the National Defense Program in the following fields:

**I. Quality Control and Specification.** The functions of a mathematical statistical nature connected with quality control and specification of articles produced by mass production are:

(1) *Tests of randomness.* These are important because statistical methods of inference are strictly valid only for random samples.

(2) *The use of probability theory in predicting the outcome of future repetitions of an operation which is in a state of statistical control.*<sup>1</sup> The evaluation of the probability that the quality of a piece of product will lie within any previously specified tolerance limits as long as a state of statistical control is maintained, and the development of sampling inspection techniques are examples of this function.

<sup>1</sup> A repetitive operation, such as a production process, is said to be in a *state of statistical control* when it produces a sequence of observations which exhibit the property *randomness*. An important aspect of quality control is the improvement of quality which comes as the result of an effort to reduce a manufacturing process to a state of statistical control. Furthermore, when this state of control is attained it is possible to gain a reduction in cost of inspection, a reduction in cost of rejections, a reduction in tolerance limits where quality measurement is indirect, and the attainment of uniform quality even though the inspection test is destructive.

(3) *Representative sampling.* When a repetitive operation such as a production process is not in a state of statistical control, it is not possible to make valid inferences about the quality of a lot from an examination of a sample from the lot unless the sampling process is one of random selection within "strata" in accordance with the principles of representative sampling.

(4) *Analysis of variance.* Reference is made here to the technique whereby the total variability of a product of an operation which is in a state of statistical control can be decomposed into components associated with the various sub-operations involved.

(5) *Correlation methods.* When a direct measurement of quality is extremely costly, it is sometimes advisable to use as an indirect measurement of quality the value of some character less costly to measure which is highly correlated with quality.

(6) *Specification of quality as a variable.* Statistical theory, including tests for randomness, must be taken into account in writing quality specifications if the consumer is to be protected against the vagaries of sampling and the producer safeguarded from the incurring of penalties of an unjust chance.

**II. Sampling Surveys.** The importance of conducting sampling surveys in accordance with the principles of *representative sampling* is well established. It is quite possible that such surveys and partial censuses will be needed in connection with the National Defense Program in order to determine the frequency and location of individuals possessing special traits, e.g. persons capable of withstanding the rigours of dive bombing, or persons possessing types of color blindness which render them valuable as observers who can detect camouflage, etc. The "problem of sizes" connected with Stores and Supplies—see below—may require careful preliminary surveys. Also, surveys may be needed to evaluate the effects of various types of propaganda.

**III. Experimentation of Various Kinds.** The mathematical statistician can be of service in connection with experimentation of various kinds undertaken as a part of the National Defense Program since the following aspects of experimentation are of a mathematical statistical nature:

(1) *Randomization.* Since statistical tests for the existence of differences between samples, of correlation, etc. are strictly valid only for random samples, the operation of randomization is of paramount importance in "the comparison of new designs, new materials or alloys, study of contact phenomena under different conditions, corrosion of materials under different atmospheric conditions, and field trial of equipment, to mention only a few." If randomization is not undertaken, observed differences between designs, for instance, may have arisen from non-random assignable differences in the material presented. Furthermore, the validity of tests for significant differences between the effects of various designs rests upon the condition that the variability observed in the effects of each design be of *random* character and free from trends and non-random shifts in magnitude—i.e. the operation of determining the effects

of each design must be in a state of statistical control, to use a phrase employed in quality control.

(2) *Experimental design.* Without careful attention to the lay-out of an experiment, the data it yields may be difficult and even impossible to interpret. Therefore, the principles of experimental design set forth by R. A. Fisher and his followers are of great importance, as are also the special experimental arrangements which have been devised to cope with many of the more usual difficulties met in practice.

**IV. Personnel Selection.** The allocation of individuals to places where they can be of greatest value in the National Defense Program will undoubtedly require tests for mental and physical traits. Although the development and analysis of such tests is largely in the hands of psychometric groups, the use of methods of multivariate statistical analysis in such work renders this field one in which mathematical statistics ought to play an important role.

It is in the above four fields that there is special need for the training and endowments of the *mathematical statistician*. He can also render valuable assistance in the following fields:

**V. Stores and Supplies.**

(1) *Problem of sizes.* Preliminary surveys are likely to prove useful in ascertaining the relative frequencies of demand for the respective sizes of clothing, etc. in different parts of the country.

(2) *Development of procedures for charting the day to day location and movement of stores and supplies.*

(3) *Problem of replacement of parts and equipment.* In many it is more economical to make replacement at statistically determined times, than to wait for complete failure.

**VI. Transportation and Communication.** Probability theory has shown its usefulness in peace time in handling "traffic" problems that arise in telephone and telegraph communication, electric power distribution, etc. No doubt it will find corresponding application to problems in these fields arising out of the National Defense Program.

**VII. Gunnery and Bombing.** Although there is a need in connection with artillery fire for further development of methods of estimating standard deviations from successive differences in order to minimize the biases arising from slowly changing conditions during the period of firing, the principles of artillery fire are quite firmly established and the relatively new science of bombing is likely to present greater opportunities for the application of the methods of mathematical statistics. For instance, in evaluating bombing techniques there is need of statistical methods in separating the constant biases from the random variability.



**VIII. Meteorology.** The extent to which statistical methods are being employed in meteorology can be seen from an examination of the Monthly Weather Review Supplement No. 39, issued April 1940, and entitled "Reports on Critical Studies of Methods of Long-Range Weather Forecasting." There seems to be excellent opportunity here for the application of methods of multivariate analysis and for the development and uses of methods applicable to serially correlated data. Such work would be of value in National Defense so far as it would enable the forecasting of conditions suitable for launching an attack.

**IX. Medicine.** The National Defense Program will probably require the preparation and storage of hormone substances, toxic compounds, drugs, and other medicinal supplies. Since many such are examined for potency, toxicity, etc. by means of animal assays, there will be considerable opportunity here for the sound application of mathematical statistics in planning and interpreting these bioassays.

In nearly all of the above activities the application of mathematical statistics is likely to encounter two major difficulties:

- (1) Obtaining an adequate trial of the methods of mathematical statistics.
- (2) Supplying persons to occupy key positions in the application of mathematical statistics in a given field—persons competent in mathematical statistics and who possess a sound background in the field of application.

In some of the above activities, e.g. Quality Control, there will be the further difficulty of

- (3) Supplying the vast number of slightly trained workers who will gather the data and perform the analyses.

It is with these difficulties in mind that the Committee recommends that the Institute

- (1) Prepare a register of Institute members, stating for each member his background, interests, and experience so far as these relate to mathematical statistics and its applications;<sup>2</sup>
- (2) Appoint a committee to handle inquiries concerning personnel qualified to deal with particular projects;
- (3) Cooperate to the fullest extent in matters pertaining to quality control and specification with the *Joint Committee for the Development of Statistical Applications in Engineering and Manufacturing*, of which the Institute is a sponsor.<sup>3</sup>

<sup>2</sup> The preparation of this register should be coordinated with any similar undertaking sponsored by the *National Roster of Scientific and Specialized Personnel*, National Resources Planning Board, Executive Office of the President, Washington, D. C.

<sup>3</sup> We suggest the following as possible undertakings in a cooperative program with the Joint Committee:

- (1) Requesting statements regarding the potential contribution to National Defense

(4) Undertake such steps as are feasible which will lead to cooperation with other organizations having interests similar to those of the Institute, e.g. the American Statistical Association, the Psychometric Society, and the Econometric Society.

(5) Establish contact with the National Defense Research Committee headed by Dr. Vannemar Bush and coordinate the Institute's activities with those of this national Committee.

In conclusion, we feel that as an organized group the Institute's primary function in relation to the National Defense Program should be to serve as a reservoir of specialists, experienced in the use of the methods of mathematical statistics, who can direct the use of these methods and be of assistance in the development of new techniques as needed. As a secondary, but equally important function, the Institute is in a position to supervise, and perhaps to undertake through the activities of its individual members, the training in mathematical statistics of the individuals who will be needed in the application of whatever statistical programs of the type noted above are undertaken in connection with the National Defense Program. *It is recommended, therefore, that the Institute's interest in the above activities, and its willingness to be called upon, be adequately publicized,* possibly by sending copies of this report to various members of the Government, such as the Chief Signal Officer and the Coordina-

of statistical methods in quality control and specification from men prominent in industry who are familiar with recent developments in quality control. Such individuals would be asked to give, where possible, concrete evidence of the value of such methods in their experience—evidence which would be helpful in securing authoritative acceptance of statistical methods in quality control and specification.

(2) The organization of a syllabus on statistical methods for use in evening courses at various industrial centers. (Captain Simon of our Committee is preparing "An Engineer's Manual of Statistical Methods" which will be issued shortly.)

(3) The preparation of a list of topics for inclusion in university courses.

(4) The preparation of a list of suggested reading on statistical methods in quality control and specification, arranged under such headings as "expository," "methodology," etc.

(5) The arrangement of local meetings and round table discussions at some of the universities in a few large industrial centers. Some well known leader of the locality might serve as chairman. To such a meeting would be invited those men in local industries who were interested in the possibility of applying statistical methods to their problems, and the meeting could be thrown open to discussion after a brief paper outlining the accomplishments of statistical methods of quality control in the speaker's experience and stating the advantages to be gained by employing such methods in the mass production of the War Preparedness Program.

(6) Sponsor the preparation of popular expository articles on quality control for industrial journals, Readers Digest, Scientific American, etc., and other activities designed to popularize the subject and gain authoritative acceptance of statistical methods of quality control.

tor of National Defense Purchases and also to the secretaries of appropriate organizations, such as the American Standards Association, with the request that they advise the Institute of any specific action they feel the Institute should take.

A. T. CRAIG

E. G. OLDS

L. E. SIMON

R. E. WAREHAM

C. EISENHART, *Chairman.*





